

Chapter 2

Literature review

The focus of the literature study was to investigate how a seamless prototype geodatabase could be developed, containing relevant existing data about the water infrastructure on the campus of the North West University, Potchefstroom. The database serves as a platform and as a source, containing relevant information for future analyses and management.

The literature review was conducted to provide a critical discussion and a brief overview of the focused GIS applications relevant to the study. For this reason many technical details about the ArcGIS geodatabase and extensions thereof are represented and discussed.

2.1. GIS, the system and the structure

GIS serves as a technological tool for understanding geography and making intelligent decisions (ESRI, 2007b). It contains sets of computer tools that allow people to work with data tied to a specific location on the earth and consists of advanced analytical mapping functions. With this increase of technological advancement, many utilities are also finding GIS and the different extensions it provides, imperative for day-to-day events. “This includes everything from automated mapping and facilities management to customer service and technician dispatch, and routing” (Harder, 1999). The capability of storing and managing large amounts of spatially related data has been enhanced through geographical information systems (Zhu et al., 2009). It also provides the ability to integrate multiple layers of information and to derive additional information (Dai, Lee & Zhang, 2001).

“GIS is a system that integrates computer hardware, software and data for capturing, managing, analyzing and displaying all forms of geographically referenced information” (ESRI, 2010a). Clarke (1986) defined GIS as “computer assisted systems for the capture, storage, retrieval, analysis and display of spatial data”. Based on common geographic locations in conjunction with the application tools that GIS provide, spatial relationships between the layers could easily be retrieved (FPA, 2010). It has proved to be a medium that provides an increase of efficiency, accuracy, productivity, communication and collaboration of spatial information. Vast amounts of spatial data could be analyzed and supported by GIS.

With the use of a standardized GIS data model, a framework could be developed in order to capture, manage and deploy spatial information. A standardized data model serves as a “living, breathing” data dictionary as explained by McLane & Yan (2009).

With the advent of Geographic Information System (GIS) in the 1970s and 1980s, cartography was changed forever (Sonnen, 2005). Unlike normal paper maps, GIS maps are interactive and adaptable. It provides the user with the ability to view, understand, question, interpret and visualize data in various ways as represented in Figure 2.1.

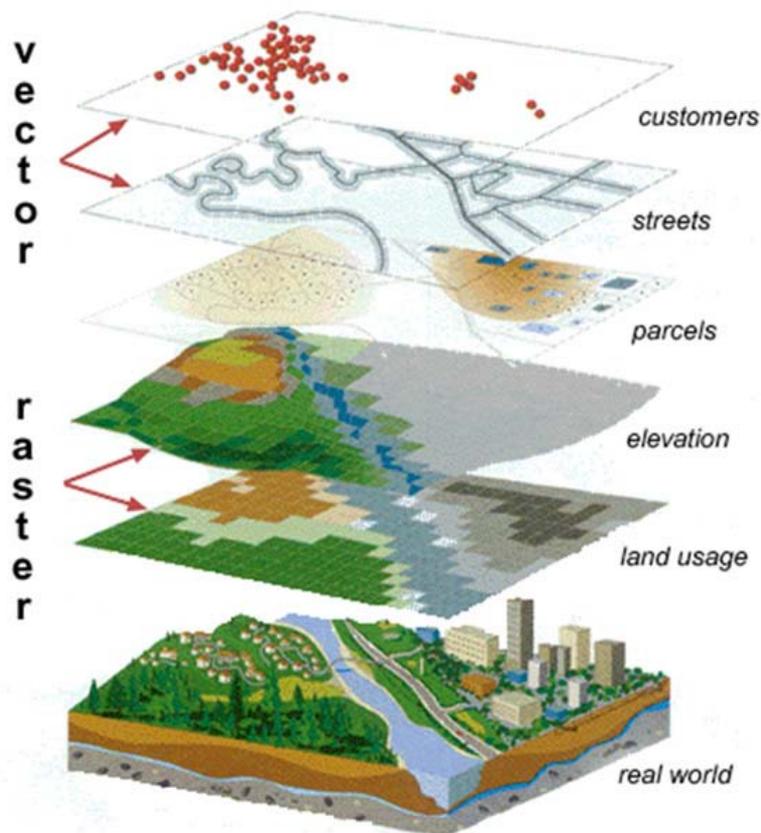


Figure 2.1. Interactive maps of GIS (Hill, 2006)

Relationships, patterns and trends in the form of maps, globes, reports and charts are represented. With the use of GIS, geographic data are organized, simplified and selectively laid out for the use of different projects or analyses regarding the earth. Each layer represents a different theme containing specific contents that defines the different layers. Features on each thematic layer are represented in the form of points, lines, polygons, rasters or tabular attributes as indicated in Figure 2.1.

GIS has the ability to overlay and match different layers of data for the same geographic area and enables one to see interactions among different datasets visually. The real world layer in Figure 2.1 is broken up in several different layers: customers, streets, parcels, elevation and land use. Each one of these layers is composed of either raster or vector data as indicated in Figure 2.1. These files are placed over each other and enable the viewer to notice correlations among these data representations. Better decision making and management could be carried out with the integrated view provided (FPA, 2010).

GIS operates using a database that is designed to work with map data (Price, 2010). Even though GIS are widely applied, a common goal and a very strong character of GIS remains the fact that it is applied to collect, manage, and analyze spatial data to produce information for better decision making (Price, 2010). Data are collected from various sources such as CAD data, as-built drawings and satellite images. Once the data have been collected it need to be saved in a central place of storage and are therefore imported into a geodatabase. The geodatabase provides functions and tools to analyze and model the data in a way that it is much more understandable.

“GIS technology integrates common database operations such as query and statistical analysis with the benefits of unique geographic examination and visualization offered by interactive maps” (ESRI, 2010a). The simplification of data offers a way that it is easily understood and easily shared. Examining the interactive maps assist the user to run relevant queries and to answer questions thereof in order to solve different geographical problems (ESRI, 2010a). As for these processes, the integration and execution of analyses on the data could not be performed if the data is not located in a central area in a certain format with a certain criteria. All the different types of data need to have a central place of storage called a database. The term geodatabase will be used as the data it contains are geographically and spatially referenced. Analyses could therefore be performed accordingly.

With the use of GIS and its geodatabases many utilities across the world are discovering an increase in the coordination and effectiveness of their overall operations (Harder, 1999). The increase of efficiency, accuracy, productivity, communication and collaboration of spatial data within many businesses or organizations have been documented to prove the success rate that GIS delivers in many departments right across the globe (Thomas & Ospina, 2004).

Many utilities around the world have made use of GIS but most of the applications fall into one of the following areas: operations, engineering, marketing, financial and mapping. The

current study and its applications mainly fall in the area of operations and engineering. When referring to operations, the focus is usually on the management and monitoring of facilities and their use.

2.2. Different types of data storage

In the following section different mediums of data storage, different types of geodatabases and a comparison between these types of geodatabases are discussed. The storage and management of geographic data in a geodatabase differs largely from other types of data storage, mainly because of the large data size that the features stored in the geodatabase consists of. Geographic data contains a vast, complicated data structure, intense operation and strong autocorrelation. In other words, geographic data could take on many diverse formats and could be represented in many more ways than that of ordinary maps. When applied in GIS it entails interactive maps describing objects and relations in space that are spatially referenced (Anon, 1996).

In the past, paper cartography has delivered paper maps that served as maps for visual representation of data and served as the “database”. The fact that the maps served a twofold purpose made the display and analysis of geographic data difficult and limited. With the use of GIS, the database, analysis and the display of data have conceptually and physically been separated into different but integrated operating systems as indicated in Figure 2.2 (Sonnen, 2005).

Referring to Figure 2.2 the transformation from paper maps towards a more centralized medium of data storage are represented. Paper maps used to be the main medium to store and represent data between the 1980s until 2002 (Sonnen, 2005). However, extensive transformation in the areas of data management and representation occurred during the year 2002 from where data within a geodatabase were applied to be represented in 2D and 3D formats (Sonnen, 2005). These representations provided the opportunity for more effective and more efficient decision-making processes and data automation. In the following section the different types of data storage and the transformation that occurred among those types are discussed.

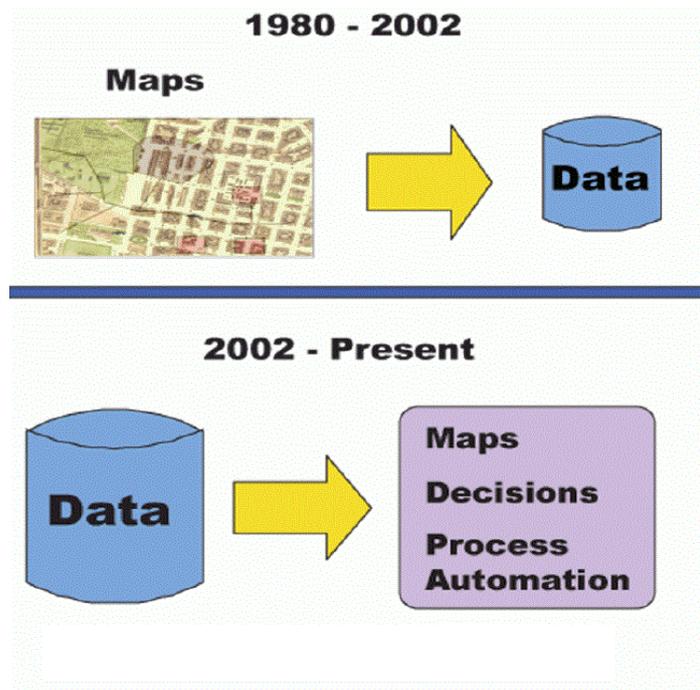


Figure 2.2. The transformation from paper maps (Sonnen, 2005)

2.2.1. Coverages

The first data models used in ArcGIS were georelational data models that stored vector data and contain both spatial (location) and attribute (descriptive) data for geographic features. These data models are known as coverages, which represents geographic features through feature classes. Each feature class contains a set of points, lines, polygons or annotation respectively through which a feature is represented. Relationships between features are determined through topology that is also contained within coverages (ESRI, 2007a).

A coverage is a spatial dataset that contains a common feature represented by the mapping of one aspect of data in space with different characteristics. According to Theobald (2001), a coverage explicitly stores topological relationships among neighbouring polygons. Coverages are also known as layers or themes (Dempsey, 2000), and are represented in many different ways. Among the examples are aerial photography, land cover data and digital elevation models. In comparison to feature data, coverage data concentrates more on spatial surroundings (Figure 2.3).

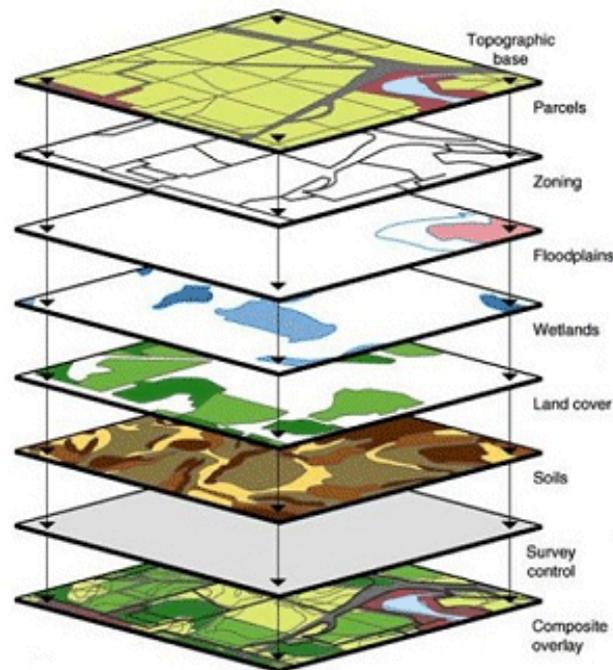


Figure 2.3. Coverages (Anon, 2012 & Poole, 2010)

As indicated in Figure 2.3 a coverage consist of different layers and each layer represents a different feature. Some key-base coverages are geo-referenced aerial photos, topographic data, soil data, and elevation models (Poole, 2010). In GIS the different layers are placed over each other to complete a composite overlay. An organized collection of coverages is called a workspace (Anon, 2012).

Coverages are managed within ArcCatalog. Referring to Figure 2.4 a coverage is stored as a directory and inside the directory each feature class is stored as a set of files, and each file contains information about a particular feature class (ESRI, 2008c). The coverage name would adopt the directory name in which the coverage is stored inside the computer.

Each coverage represents a different feature set such as streams or roads as indicated in Figure 2.4. These features consists of different coverage feature classes which may contain points, lines (arcs), polygons, annotation and tic files. The features in a coverage are usually defined by more than one feature class also indicated in Figure 2.5. Tic points are feature classes that are part of every coverage because it defines the extent of a coverage and represents known real-world co-ordinates (ESRI, 2007a).

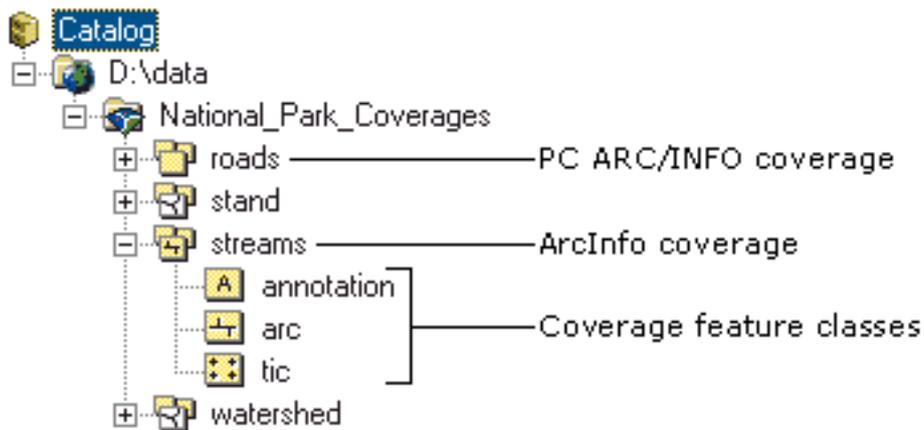


Figure 2.4. The appearance of the coverage in an ArcCatalog directory (ESRI, 2007a)

Coverages store location and shape of geographic features in a very accurate way according to resolution and precision. However, external factors such as input data sources and the tools used for the input of data could influence the resolution of the coverage. These factors are:

- the co-ordinate precision specified;
- the precision of the input device; and
- the scale of the input documents and coverage tolerance.

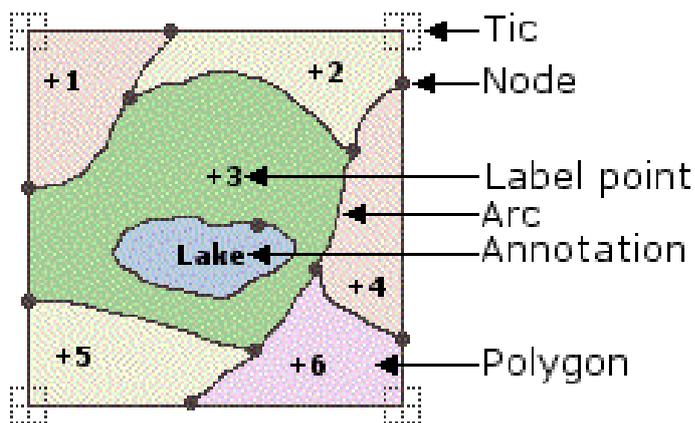


Figure 2.5. A detailed view of the coverage (ESRI, 2007a)

2.2.2. Shapefiles

With the release of ArcView 2 early in 1999, the introduction of shapefiles made their occurrence in GIS (Theobald, 2001). The shapefile format is created by ArcView and can be used by ArcView, ARC/INFO and other widely used GIS software that are used for making maps and analyzing geographic data (USGS, 2010).

A shapefile is known as a digital, non-topological vector storage format, which stores associated attribute information and geometric location. It also stores geographic features such as points, lines, and polygons with attribute data as a collection of files. In order to prevent deactivation of the data, the files need to be moved as a group (Figure 2.6).

Spatial relationship information such as connectivity, adjacency, and area definition are not maintained by a shapefile since it is non-topological. Even though shapefiles are non-topological of nature it serves as an essential component of GIS software and is commonly applied to import and export data to and from it (Hijmans et al. 2001).

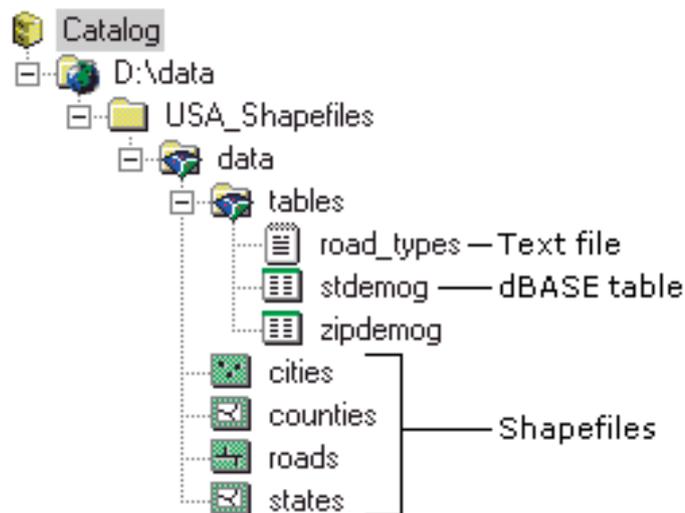


Figure 2.6. Representation of the shapefiles in ArcCatalog (ESRI, 2008a)

Using a shapefile format is therefore much simpler but less capable when performing complex spatial analysis (USGS, 2010). Shapefiles are managed in ArcCatalog and editing them could take place with any license level in ArcGIS. Overall, they have faster drawing speed and edit ability and require less storing space. Whenever more advanced editing such as topology need to be applied to the data, the shapefiles first need to be imported into a geodatabase. When it is imported the feature types in the shapefile are automatically

converted to geometry types in a geodatabase. Shapefile feature types have an advantage over coverages whenever they are imported. Shapefiles are much more similar to the geometry types stored in a geodatabase that makes the data conversion much simpler than those of coverages (ESRI, 2008a).

2.2.3. Feature classes

Feature classes have recently been applied to serve as the new mediums of data storage within the geodatabase. Feature classes are generally described as a collection of geographic features that consists of the same geometric type and spatial representation such as points, lines, polygons, multipoints, annotation, dimension or multipatch (University of Alberta, 2010). Within a feature class each individual feature – point, line, polygon – are represented as a separate object with great versatility within the geodatabase (Rich et al, 2002).

With reference to Table 2.1 feature classes share a common set of attributes. This implies that a feature class is a collection of homogeneous features representing the same geographic elements such as manholes, valves, mains and pumps.

Grouping different feature types together this way provides the ability to process them as a single unit. Inside feature classes, well defined integrity rules maintains data integrity and ensures that individual features can share spatial relationships with other features. Additional properties could also be specified inside feature classes that promote their functionality (Arctur & Zeiler, 2004). A brief summary of the three data storage types have been provided in Table 2.1. Comparing the three types with each other accentuates the differences between those formats. Even though shapefiles and coverages are still used, the nature and advantages that feature classes provide within the advanced operating capabilities of the geodatabase motivates its great operating functionality.

Feature classes could be stand alone or they could be grouped as a collection of related features inside another medium called a feature dataset. However, a feature class cannot exist outside a geodatabase but surely outside a feature dataset (Figure 2.7).

	<i>Coverages</i>	<i>Shapefiles</i>	<i>Feature classes</i>
Data source	<ul style="list-style-type: none"> • First data models • Store vector data • Contains spatial and attribute data of geographic features 	<ul style="list-style-type: none"> • Early 1999 • Digital non-topological vector data storage format • Contains attribute data as collection of files • Stored as a set of related files 	<ul style="list-style-type: none"> • Newest medium of storing and organizing spatial data inside a geodatabase • Topological data storage
Representation	<ul style="list-style-type: none"> • Geographic features represented through feature classes (consist of points, lines and polygons) • Each coverage represents a different feature set 	<ul style="list-style-type: none"> • Geographic features represented through points, lines and polygons 	<ul style="list-style-type: none"> • A collection of geographic features consisting of the same geometric type and spatial representation such as points, lines and polygons • Each individual feature are represented as a separate object • Feature classes also share a common set of attributes
Relationships	<ul style="list-style-type: none"> • Spatial relationship information are determined through topology 	<ul style="list-style-type: none"> • Spatial information are not supported by topology 	<ul style="list-style-type: none"> • Well defined integrity rules ensures that individual features can share spatial relationships – supports topology
Advantages	<ul style="list-style-type: none"> • Store shape and location of geographic features in a very accurate way • Uses a simple structure to maintain topology 	<ul style="list-style-type: none"> • Editing could take place with any license level • Overall faster drawing speed and edit ability, and requires less storage space • Much more similar to the geometry types stored in a geodatabase which makes the data conversion much simpler than those of coverages 	<ul style="list-style-type: none"> • Feature classes are very versatile and could be contained within a geodatabase of a features dataset • Features could be processed as a single unit through a feature class • Additional properties could also be specified inside feature classes which promote their functionality
Disadvantages	<ul style="list-style-type: none"> • Coverage data is generally mainly focussed on spatial surroundings than on feature data 	<ul style="list-style-type: none"> • Using a shapefile format is much more simple but less capable when performing complex spatial analysis 	<ul style="list-style-type: none"> • Can only exist inside a geodatabase

Table 2.1. A compared summary of the three data storage types (ESRI, 2007a; Theobald, 2001; FPA, 2010)

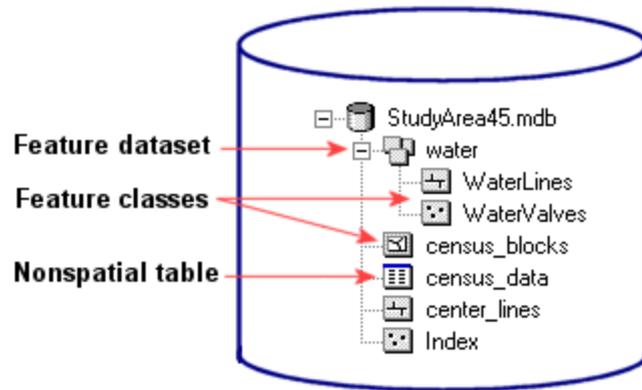


Figure 2.7. A geodatabase representation containing feature classes and feature datasets (ESRI, 2002c)

2.2.4. Feature datasets

Feature datasets serves as the containers for data that are spatially related along with relationship classes, geometric networks and geodatabase topologies (Arctur & Zeiler, 2004). Data are stored thematically in a logical grouping of different feature classes inside a feature dataset. Every feature or layer inside the feature dataset has a defined spatial reference. This way connectivity and topology of features that touch, coincide, overlap, cover and intersect each other are enforced. Spatial relationships among related feature classes are managed with its location inside a feature dataset. This ensures that simple stand-alone feature classes and other more advanced collections of features operate as a system of objects and relationships.

2.2.5. Geodatabases

Geographic information systems have been the technological advanced medium with the capability to organize information into a series of layers that can be integrated using geographic location. GIS data storage, on a basic level, operates through the storage of data inside a geodatabase. Spatial (geographic) data are stored and organized into datasets as a series of thematic layers to represent and answer questions about a particular spatial problem. Harder (1999) describes GIS as data being broken up into several different parts and organized as a set of layers or themes all related by location and stored as a unity inside a geodatabase. With reference to Figure 2.1 reality is sliced into layers and each layer graphically represents a different feature such as water facilities, water features, water lines and equipment, hydrology, tax parcel management, transportation or environment, fire

locations, buildings, orthophoto imagery, and raster-based digital elevation models (DEMs) (Arctur & Zeiler, 2004).

All the different features of the layers are stored separately but contain co-ordinates that allow the software to draw or place the feature in the correct location with regard to the earth and related features. Once referenced the features could operate as a utility information model. The basic requirement of such a utility information model is to ensure that all the parts of the system operate well at all times. To simplify this procedure the components of the system need to be easily retrieved from a database when the features within the database are spatially referenced.

Information that can be represented through datasets is:

- raw measurements such as satellite imagery;
- compiled and interpreted information such as utility records and CAD data; and
- data received during geoprocessing operations for analysis and modelling.

A geodatabase is therefore defined as a geographic database and stores geographic information inside a database management system (DBMS) (CSISS, 2010). Harrison et al. (1990) describes a database management system as “software that allows one or many persons to use and/or modify the data inside a database”. According to Price (2010), a geodatabase serves as a container to which feature classes and other database objects could later be added.

The emphasis of the geodatabase design is placed on identifying the thematic layers that need to be used, specifying the contents and representations of each thematic layer such as attributes, relationships between attributes and relationships between features. The geodatabase not only defines how data is stored, managed and accessed, but also provides the users with the ability to maintain a consistent, accurate geospatial database as well as implementing complex business logic (Law, 2007). The geodatabase organize data into feature classes, attributes and relationships and provides the possibility to add spatial integrity rules for the data such as topology, relationship classes and geometric networks (Arctur & Zeiler, 2004).

Many more functions are presented by a geodatabase that mainly serves as a centralized medium of storage for a wide variety geospatial information in a DBMS. Multiple formats of spatial data are supported and according to Law (2007) include sources such as:

- simple features such as shapefiles and coverages;
- custom features with business logic and editing rules;
- attribute data;
- metadata;
- images;
- raster/grid data; and
- CAD data.

Geodatabase provides many advantages over other data storage types such as shapefiles and coverages. These advantages are:

- a geodatabase could store multiple feature classes in the same file and therefore makes it much more superior than the shapefiles;
- related feature classes with the same spatial reference could be grouped under one directory named a feature dataset;
- predefined specifications could be added for each field in a geodatabase feature class or table; these specifications are known as domains;
- labels are saved to annotation feature classes in the geodatabase;
- advanced capabilities such as geometric and logical networks, true curves, complex polylines, and user defined features are all supported by a geodatabase;
- large collections of objects in a database table and geometrical features are supported by a geodatabase (CSISS, 2010); and
- relationships between objects could be established through the creating of relationship classes inside the geodatabase (CSISS, 2010).

More advanced functionalities are also available when working with multi-user geodatabases or with file geodatabases in ArcEditor or ArcInfo. These advantages are:

- relationship classes are created between feature classes in feature datasets. For example, when two features such as water valves and water lines are connected to

each other, a relationship class between the two ensures that the water valves will move along with the water mains whenever it is moved;

- geometric networks and network datasets are created in order to model connectivity and to perform trace and path analysis;
- different versions of data are stored; and
- custom features are stored. These are features that represent real world features more accurately (ESRI, 2008).

Data storage could be maintained in many different ways and in many different types of databases. Computer experts usually focus more on the design of commercial databases, while geographic information experts apply their minds to mediums of geographic data processing.

The expression of a geodatabase design is normally done by means of a data model. “Data models are sets of concepts describing a simplification of reality expressed in database structures such as tables and relationships and they provide standardized frameworks for users to store information and serve as the basis for applications” (Strassberg, 2005). With a series of steps and assumptions, raw data of a project are converted into an organized set of useful information. Strassberg (2005) stated that a geodatabase is also known as a spatial database that stores spatial database structures and are applied to describe geospatial phenomena.

2.2.5.1. Types of geodatabases

The geodatabase provides three levels of expandable geodatabases. The first level is scalable (expandable) geodatabases that includes enterprise edition, workgroup edition, and personal edition geodatabases. The second and third level geodatabases are known as the additional geodatabases and includes file geodatabases and personal geodatabases. The most appropriate type of geodatabase for a specific purpose would depend on specific requirements for the project at hand (Law, 2007).

Referring to Table 2.2 a comparison between the different databases with regard to their storage formats, storage capacity, setup and management, supported operating system, number of users, what it is designed for, and additional information, have been created. In order to obtain a good understanding of all the different types of geodatabases and finally to

	Additional geodatabases		Scalable geodatabases (three levels)		
	Personal geodatabase	File geodatabase	Enterprise edition	Workgroup edition	Personal edition
Storage format	Microsoft Access	File folders of two composed files	DBMS	DBMS	DBMS
Storage capacity	2 GB	1 TB = fast data storage	Range in different sizes (can be scaled to any size)	4 GB Max limit	4 GB Max limit
Setup and management	ArcCatalog	ArcCatalog	Database administrator (requires advanced administration skills and knowledge)	ArcCatalog	ArcCatalog
Supported c/s platform	Windows	Windows and Unix (Solaris and Linux)	A range of operating systems – IBM DB2, IBM Informix, Oracle, Microsoft SQL Server. Requires ArcSDE data server software	ArcGIS Server Workgroup and Microsoft SQL Server 2005 Express. Requires ArcSDE data server	Microsoft SQL Server 2005 Express – available with ArcInfo and ArcEditor. Requires ArcSDE data server
Number of users	Single editor, multiple readers	Single editor, multiple readers	Multiple editors and readers, 24/7, on computers of any size	10 simultaneous users (users could be editors)	3 simultaneous users (1 editor, 2 viewers)
Designed for	Single user working with small datasets	New users and local use	Large scale scenarios	Small mid-size department applications	Disconnected editing and application scenarios
Additional information	Available since release of ArcGIS, does not support all the features of the full gdb. Limited scalability and functionality	Rich structure – users could manage and model data easily. No versioning or long transactions. More scalable and better functionality than personal gdb. Vector data could be compressed	Larger edition	Smaller edition	Smallest edition. Provides full geodatabase support to Arc Editor and ArcInfo

Table 2.2. A summarized representation of the different databases (Detwiler, 2002; Geographic Technologies Incorporated, 1995; Law, 2007)

The databases available in GIS are divided into two sections; additional geodatabases and scalable geodatabases. Additional geodatabases are divided in personal and file geodatabases and are used for personal use or projects operating with smaller data storage than those of the scalable geodatabases. The scalable geodatabases are divided in enterprise, workgroup and personal geodatabases and are generally server specific, applied for larger projects and enterprise applications for data storage in various departments. It could also be used by more than one user at a time in comparison to the additional geodatabases that could only be used by a single user. The scalable geodatabases could also be managed with a range of operating systems as to where the additional geodatabases have a focused operating system, varying between Windows and Unix as indicated in Table 2.2.

2.2.5.2. ArcSDE geodatabases

ArcSDE geodatabases are known as the three scalable geodatabases - ArcSDE Personal, ArcSDE Workgroup and ArcSDE Enterprise geodatabases - because they operate in association with the ArcSDE application server that facilitates the storing and managing of spatial data. IBM DB2, Informix, Microsoft SQL Server, SQL Server express and Oracle are some of the commercial databases also known as relational database management system (RDBMS), which stores spatial data. Also, according to Geographic Technologies Incorporated (1995), “ArcSDE serves as the mediator between GIS clients and the RDBMS” (Figure 2.8).

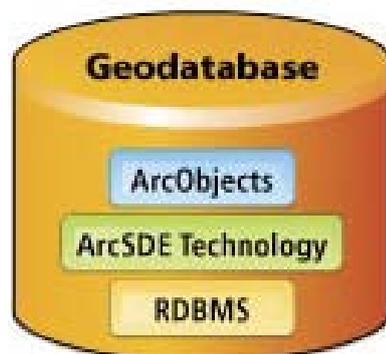


Figure 2.8. Representation of the ArcSDE geodatabase (Geographic Technologies Incorporated, 1995)

In basic terms, ArcSDE technology is data server software that serves as an advanced technological medium and enables the user to easily store, access and manage spatial data in a relational database management system (RDBMS). Organizing and analyzing large amounts of diverse geospatial data is the problem that many engineering firms and businesses face (McLane & Yan, 2009). With reference to the aforementioned section, this problem could be solved with the application of a geodatabase.

2.2.5.3. Enterprise GIS

Enterprise GIS provides the opportunity for more than one user to access the same data in multiple ways. Traditional software programmes, Web browser applications and wireless mobile devices are among the ways to access data. The needs of many different users are met through the shared data access. Rich, Das & Kroot (2002), describes an enterprise in the context of GIS as, “Any organization that needs to support multiple simultaneous users accessing a shared information resource” (Figure 2.9). According to the latter, an Enterprise GIS are also defined as a common spatial database is that is applied in various areas and departments right across the globe. This may include many people, even thousands of people networking together but it could also be as little as four people working on a single project. Large amounts of simultaneous users can perform queries, access resulting data, perform analyses quickly and easily and execute tasks without having to submit requests to outside departments. With the ArcGIS Server as a supportive medium, the internal processes of an enterprise GIS environment are improved (ESRI, 2007b). The necessity of a centrally managed database with secure, dependable access is imperative for the functionality of an enterprise GIS system and also requires a Relational Database Management System (RDBMS). Applications such as security, record level locking, editing of conflict resolutions etc. enhances the requirements of the RDBMS inside the database (Rich, Das & Kroot, 2002).

The advantages of implementing an enterprise GIS is that it saves time, provides direct access to data and frees up GIS analysts so they can perform more technical, GIS-centric work for an organization. GIS data processes such as geocoding, mapping, geoprocessing and data management could also be combined with other complementary enterprise systems or enterprise resource planning by means of an integration platform between ArcGIS Desktop software and ArcGIS Server (ESRI, 2007b).

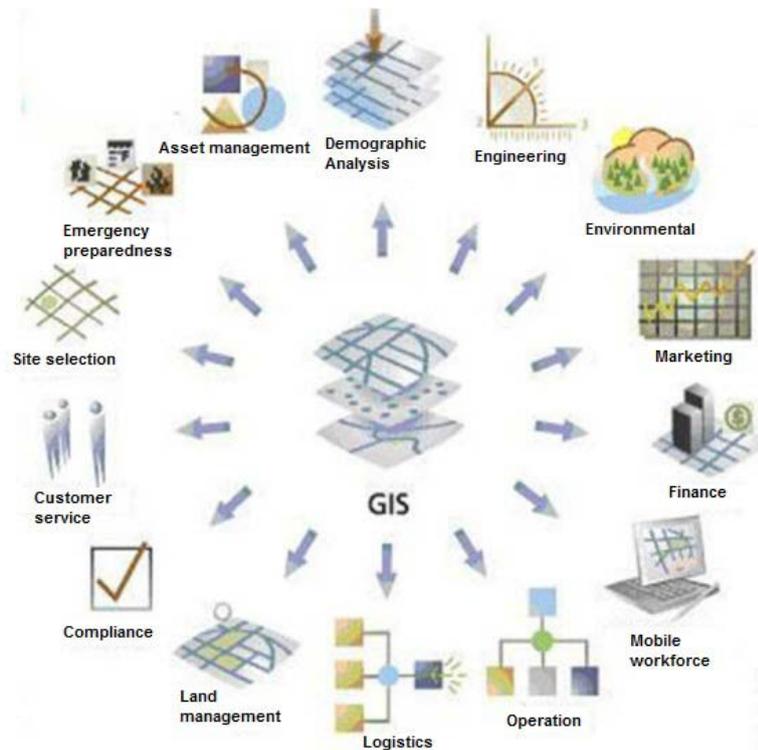


Figure 2.9. Enterprise GIS (Support Systems Limited, 2009)

Concluding the section, based on the literature, it was found that the geodatabase is a sufficient medium to store and manage geographical information in an organized way. The question however would be what geodatabase would be best suited for the project at hand? According to Table 2.2 there are geodatabases for every type of application. According to an article by Childs (2009), due to the structural performance and data management advantages over personal geodatabases and shapefiles, the file geodatabase could be applied for any size dataset, large or small. The file geodatabase could also be used for any project, whether it is a single-use project or a project involving a small group with multiple editors. Referring to Table 2.3, Childs (2009) divided the advantages of the geodatabase in three sections. These sections motivate the application of the geodatabase and support the information provided in Table 2.2. According to the advantages indicated about the file geodatabase it is definitely the most suitable geodatabase for the study on the campus.

The advantages of the file geodatabase	
Structural	<ol style="list-style-type: none"> 1. Improved versatility and usability 2. Optimized performance 3. Few size limitations
Performance	<ol style="list-style-type: none"> 4. Easy data migration 5. Improved editing model 6. Storing rasters in the geodatabase
Data management	<ol style="list-style-type: none"> 7. Customizable storage configuration 8. Allows updates for spatial indexes 9. Allows the use of data compression

Table 2.3. The advantages of the file geodatabase in three sections (Childs, 2009)

2.3. Geodatabase design

The geodatabase forms the core of the GIS system and the functionalities thereof have already been mentioned. Nonetheless, it is also important to look deeper into the structure of the geodatabase, with the focus on the design principles and different components inside of it that gives the geodatabase the advantages it is known for.

2.3.1. Representations

Geographic representations are one of the essential components of a geodatabase. These geographic representations are composed of geographic entities and the way they are represented can be divided into three categories (ESRI, 2010a):

- features (points, lines and polygons) representing vector data;
- continuous surfaces and imagery using rasters and triangulated irregular networks (TINs); and
- map graphics such as text labels and symbols.

2.3.2. Thematic layers

As mentioned earlier, all the different features represented inside a geodatabase are organized into a set of thematic layers, each containing relevant information as a collection of common geographic elements (refer to Figure 2.1).

The motivation behind these thematic layers originated from the requirement to organize the information in a logical and representable format. Each thematic layer is defined by means of attribute information and specified properties. For example, a few businesses in an area could be represented by a few points in a layer to which descriptive information about each business could be assigned respectively. Each thematic layer could also be managed independently of each other serving as independent information sets. Spatial reference is also defined in each layer that provides the ability for the layers to be placed over each other (CSISS, 2010).

2.4. Inside a geodatabase, the structure and the design

As previously mentioned by Price (2010), a geodatabase in its most basic form, serves as a container with the ability to store and manage data. According to the Environmental Systems Research Institute (ESRI), the geodatabase and the ArcInfo coverage are very alike. As described by Price (2010), both geodatabases and coverages can store topological relationships. One major difference between the two storage types, is that the geodatabase is much simpler to construct and more robust for general use.

According to Arctur & Zeiler (2004), every geodatabase comprises of a database schema which enables the information inside a geodatabase to be represented as thematic layers. This database schema includes definitions, integrity rules, and predefined data behaviour for an integrated collection of datasets. The database schema is subjected to operate in coordination with the core elements of the geodatabase. The core elements of the geodatabase includes feature classes and feature datasets, topologies and networks, raster datasets and raster catalogs and also contains properties to define each one of these elements. The database schema, together with the core elements of the geodatabase, forms the essence of the design and structure of the geodatabase. Another advantage of the geodatabase is that it provides the option for the user to design a geodatabase together with the description of all the objects in it without any actual data inside the database. This means that the created geodatabase could be used to generate several geodatabases with the same structure (Price, 2010).

A geodatabase, whether it is a personal or multi-user, is therefore a storage mechanism for spatial and attribute data, and provides a storage structure for the features, collection of features, tables, geometric networks, relationships between attributes and relationships between features. ArcCatalog, ArcToolbox and ArcMap serve as tools through which the geodatabase could be created, accessed and managed.

2.4.1. Important geodatabase design elements

The contents of a geodatabase design are represented by the core elements. In connection with the previous section about the database schema, the key elements are datasets, relationship classes, domains, spatial relationships and spatial rules, map layers, and 2D and 3D base maps. These key elements are very helpful for the way data inside the geodatabase is documented. It is therefore important to provide a good overview of the following elements as they will serve as the “cement” that will bind the “bricks” (features and entities), in the geodatabase together. In this section there will be focused on: datasets, relationship classes, subtypes and domains, topology and geometric networks.

2.4.1.1. Datasets

The datasets in the geodatabase provides the possibility for the user to add different descriptive specifications to attribute tables, feature classes, and raster datasets in the geodatabase (Arctur & Zeiler, 2010). Each one of these elements could have properties defined and recorded for it in order to ensure data accuracy and consistency. Table 2.4 represents the storage format for features and their attributes within the geodatabase. These data type formats could be any of the following;

Numeric data types	<ul style="list-style-type: none"> • short integers (a two byte number, between -32,000 and +32,000, no decimal numbers); • long integers (a four byte number, between -2 billion and +2 billion, no decimal numbers); • floats (a four byte, single-precision floating point numbers, for decimals); and • doubles (an eight byte, double-precision floating point numbers, for decimals)
Text fields	<ul style="list-style-type: none"> • alphanumeric symbols, using text and numbers, for example assigning a number 1 representing plastic PVC pipes and a number 2 representing copper pipes.
Date fields	<ul style="list-style-type: none"> • for storing dates, times or both
BLOB fields	<ul style="list-style-type: none"> • in the geodatabase these represent a long sequence of binary (two) numbers

Table 2.4. The storage format for objects within the attribute table (ESRI, 2005a)

2.4.1.2. Relationship classes

Inside a geodatabase some features have associations with each other that are termed relationships. Rows in one table are associated with rows in another table. For example a pipe fitting may be located in a certain space type such as an office or a classroom. However, the relationships inside a geodatabase between different objects are not limited only to spatial objects such as features, stored in feature classes. Relationships could also exist between non-spatial objects, stored in the rows of a table. These relationships are stored in relationship tables inside a geodatabase.

There are three basic types of relationships that could be established inside a geodatabase, and these are one-to-one, one-to-many, and many-to-many cardinalities as indicated in Figure 2.10.

Whenever a relationship between two objects is established, they are maintained through attribute values for key fields. The key fields are those fields that define to which object a

particular feature in one table are related to in another table. The primary key field is the field that serves as the originating field. The foreign key field is the field that serves as the destination field. Whenever a relationship between two fields is established, the tables take on each other's attributes. This means that the two features are connected by means of a corresponding field that contains attributes mutual to both features. In case of a one-to-many relationship, the feature in the origin feature class is related to many features in the destination feature class.

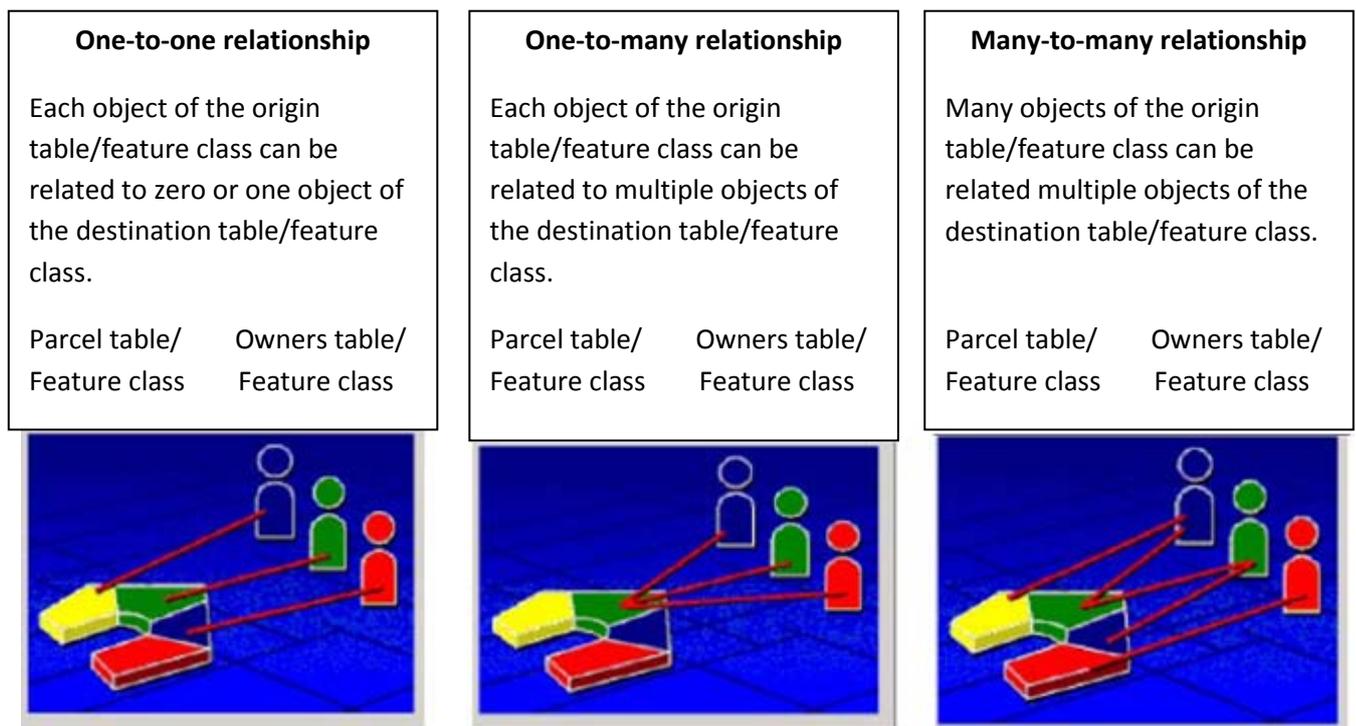


Figure 2.10. Three types of relationship classes (ESRI, 2005c)

In case of a one-to-one relationship, the origin feature has a relationship with only one feature in the destination feature class. Whenever there are more than one feature in the origin feature class that are related to more than one feature in the destination feature class, a many-to-many relationship is established (ESRI, 2005c).

Relationship classes contain path labels that serve as descriptions that indicate the direction of relationship connection. It is similar to a two way road; the forward relationship indicates the navigation from the origin to the destination and vice versa. With the establishment of a few relationships, the data consistency between the different objects inside the geodatabase is enhanced. For example, if a piece of the pipeline is shifted to another location, the valve will shift with the pipeline. However, there are relationships that could exist independently of

each other. They are called “simple relationships” (ESRI, 2005c). Composite relationships are established between two or more features that are dependent on each other.

Figure 2.11., indicates that each parcel has a related Parcel_ID in the “Permit” table. The primary key is the “Parcel_ID” in the “Parcel” table. Parcel_ID field in the “Permit” table serves as the foreign key. In this instance a relationship between spatial and non-spatial objects are established. It is also a many-to-many relationship because one parcel could have many permits and one permit could be handed out to more than one parcel (ESRI, 2010b).

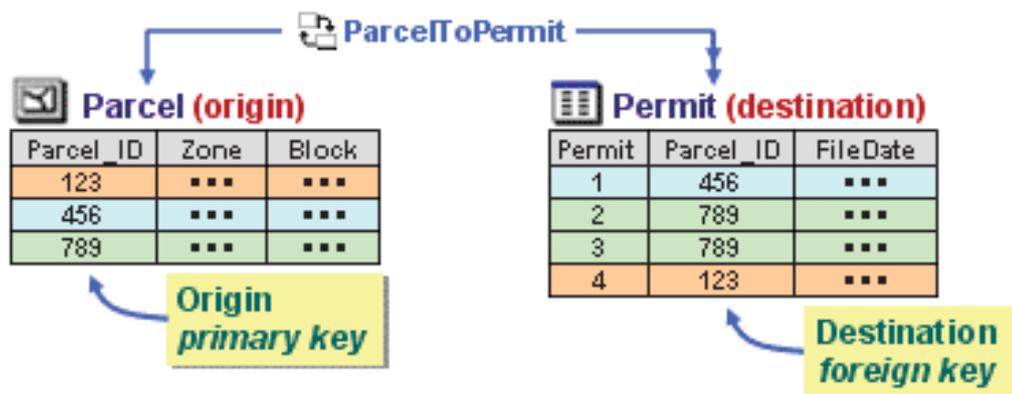


Figure 2.11. Relationships between parcels (ESRI, 2005c)

2.4.1.3. Subtypes and domains

The basic concept of subtypes and domains are to ensure that data entry and specifications inside the system being modelled are accurate and consistent. In other words, it ensures data integrity.

Subtypes and validation rules

Subtypes are defined as a subclass of features in a feature class or objects in a table that share the same attributes. Subtypes provide the possibility to categorize and organize data without having to split the data into separate layers (Taggart & Ridland, 2000). This in return ensures automatic data entry resulting in data consistency in the geodatabase. An example of subtypes is the material type for pressurized water mains. There are different types of materials the mains could be composed of: cast iron, ductile iron, or copper. Referring to the attributes, these pipes are “pressurized mains”, they could only have certain sizes and ground surface types.

Subtypes offer many advantages for the geodatabase in areas of data integrity and performance. It decreases the amount of feature classes that need to be created inside the geodatabase because a subtype could be created in the place of a feature class (Mandloi, 2007). A default value for each subtype could also be created and could be applied whenever a feature is created. In order to ensure the editing of valid sets of information, coded or ranged domains could be created for the different subtypes. Subtypes only permit long integer data fields whenever created inside an attribute table (Taggart & Ridland, 2000). Connectivity rules could also be established between different subtypes and feature classes. Creating topology and relationship rules between subtypes, tables and feature classes ensures correct connectivity between features (FPA, 2010).

With reference to the advantages of the previous section, subtypes enhance time efficiency and simplify the data entry process of attribute values. Referring to Figure 2.12, each subtype description consists of a subtype code value by which an object's subtype is determined.

Subtypes of PressurizedMain			List of defined default values and domains for subtypes in this class		
Subtype field		Subtype			
Default subtype		1			
Subtype Code	Subtype Description	Field name	Default value	Domain	
1	BlowOff				
2	Bypass	Diameter	10	MainDistributionDiameter	
3	AirRelease	JointType1	Flange	JointType	
4	ChemicalInjection	JointType2	Flange	JointType	
5	DistributionMain	WaterType	Potable	WaterType	
6	Interconnect	Material	PVC	WaterLineMaterial	
7	PipeBridge	AdministrativeArea	N	AdministrativeArea	
8	SamplingStation	OperationalArea	D1	OperatingArea	
9	TransmissionMain	LifecycleStatus	Active	LifecycleStatus	
		Enabled	1	Enabled	

All subtypes share the same default value and domain values

Figure 2.12. Subtypes of the "PressurizedMain" (ESRI, 2011c)

Each subtype table has defined default values and domains for the subtypes in a certain class as represented on the right hand of the table. These include field names that correspond with the field names in the attribute table, default values, and domains for each field. Different connectivity rules are also associated with it.

When to use subtypes

The question, however, arises; when should subtypes be applied and when should additional feature classes be created? As explained by ESRI (2005a), whenever objects should be distinguished by their default values, attribute domains, connectivity rules and relationship rules, subtypes could be applied. Controversially, whenever objects need to be distinguished based on their different behaviours, attributes, access privileges or when the objects are multiversions, it is more suitable to create additional feature classes. Subtypes could also be used whenever attributed features already exist and consistency needs to be maintained (Taggart & Ridland, 2000).

Attribute domains

Applying attribute domains inside a geodatabase ensures that the software maintains data integrity in certain attribute columns (fields). It represents a list of valid predefined values for attribute columns that allows automatic data entry (Taggart & Ridland, 2000). Attribute domains are defined at the creation of the geodatabase, as well as throughout its use, with values that could be used for any features in the geodatabase and could be shared among different feature classes and tables in the geodatabase. Attribute domains are predefined which means that they serve as a restriction to allowable values in any particular attribute field for a table, feature class, or subtype. It is therefore another way to ensure ultimate data integrity in the geodatabase which means that incorrect editing of features are eliminated (Taggart & Ridland, 2000).

Whenever a new feature needs to be created and added to the geodatabase, the predefined domains would be available by means of a dropdown box provided with the predefined options to choose from. An example would be the possibility of editing an attribute such as “Maine” instead of “Main” or by entering a number instead of a name. With predefined attributes problems such as these could be prevented. It is very important to ensure that the data type of the participating field corresponds with the data type of the domain that should be used in that particular field. A domain would not be available for a field if the data type of that field differs from the data type of the domain.

Range and coded domains

Two types of attribute domains exist: range domains and coded value domains. Range domains provide a range of values for entities with numeric data whenever a range of values

need to be specified. For example, a distribution main may have a certain range of pressures it could handle, between 40 and 100 psi. In this instance a range domain could be specified for this range in pressures (Arctur & Zeiler, 2004).

In the case of coded value domains, valid values are assigned to features with values that have specific codes. This includes attributes for text, numbers, dates, etc. An example of domains would be the different material types a water main could be created from. The material types could be: CI (Cast Iron), DI (Ductile Iron), PVC (PVC), ACP (Asbestos concrete), COP (Copper). Whenever a new pipe or feature is added to the model, one of these materials could be selected from the list in order to define the specific material type of the main without the need to type it over for each pipe section (Arctur & Zeiler, 2004).

Split and merge policies

Attribute domains have another function for defining features. These are split and merge policies. Whenever a feature is divided during an operation such as the splitting of a main pipeline, for example to channel the flow of water into two directions, the separation is controlled by the split policy. If one pipe is merged with a different pipe into a single feature, the attributes are controlled by the merge policy.

Each of these policies has three additional policies for the attributes of a feature in any given table, feature class, or subtype and are indicated in Table 2.5.

Split policies		Merge policies	
Default value	Attributes of two resulting features takes on the default value for the attribute of the given feature class subtype	Default value	Attributes of two resulting features takes on the default value for the attribute of the given feature class subtype
Duplicate	Attribute of the resulting features takes on a copy of the original objects attribute value	Sum values	Attribute of the resulting feature takes on the sum of the values from the original features' attribute.
Geometry ratio	The attributes of the resulting features takes on the ratio in which the original geometry is divided	Geometry weighted	Attribute of the resulting feature is the weighted average of the values of the attribute from the original features.

Table 2.5. Split and merge policies (ESRI, 2005a)

2.4.1.4. Topology

In order to ensure data integrity in GIS datasets, such as accurate and organized connection of points, lines and polygons, topology and topology rules are examined. According to McDonnell & Kemp (1995), the study regarding properties of objects including adjacency, connectivity and containment defines in part the term topology. It is one of the earliest identified instances used for problem solving in the Königsberg Bridge problem in 1736. “GIS Topology is widely known as a set of integrity rules that define behaviour of how points, lines and polygons share geometry” (ESRI, 2005b). Referring to Price (2010), “a topological data model not only stores features, it also contains information about how the features are spatially related to each other.”

Topology operates by means of integrity rules that enable the user to model spatial relationships. Features could be related and could interact in many ways, for example, whether two parcels of land might share the same common boundary (adjacency), whether two water lines might be connected to each other (connectivity), whether two features might overlap each other such as state boundaries that are incorrectly subdivided (overlap), or in the event of water pipelines that connects to a water main or two roads intersecting each other.

The accuracy in which real-world relationships between features are represented is also evaluated by topology. This capability evaluates the logical consistency of features. Examples thereof, includes situations such as the definite connection of two pipelines when represented by two lines that connect or to prevent a line or polygon boundary to cross over itself.

Topology rules such as “Lots cannot overlap one another, Lots must completely cover Parcels, Lot Lines boundaries must be covered by Parcels”, are some of the rules to ensure that data are correctly maintained and that advanced feature behaviour and integrity rules are added (Figure 2.14). Referring to Figure 2.13 six new topology rules are provided by ESRI with the release of ArcGIS 10. The rest of the 36 topology rules are available from ESRI’s website (ESRI, 2002b).

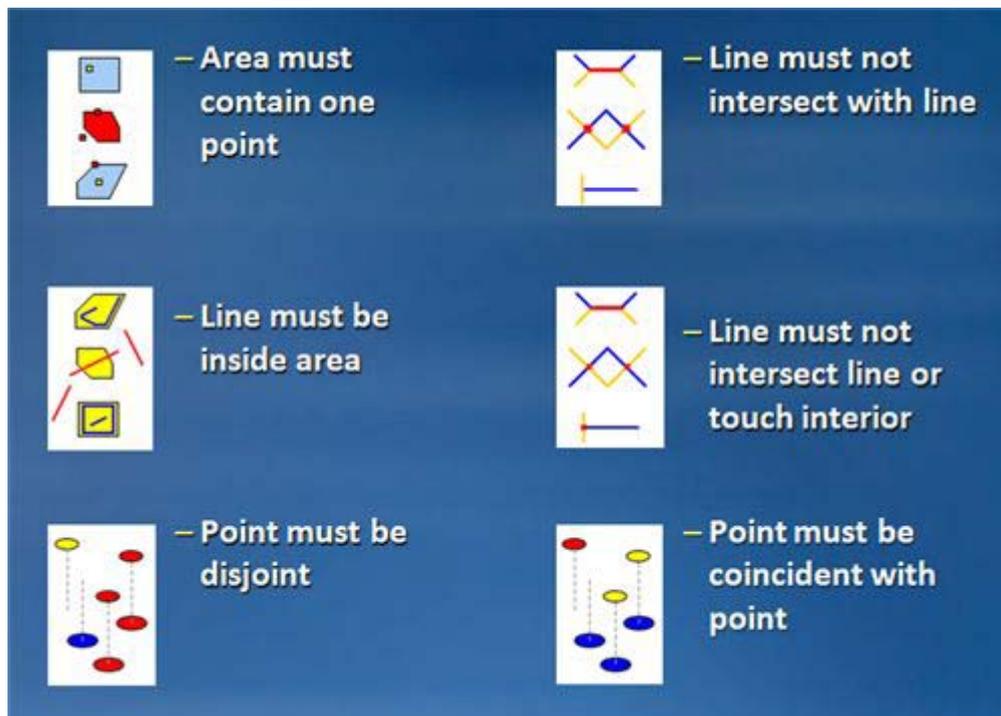


Figure 2.13. Six new topology rules with ArcGIS 10 (ESRI, 2011a)

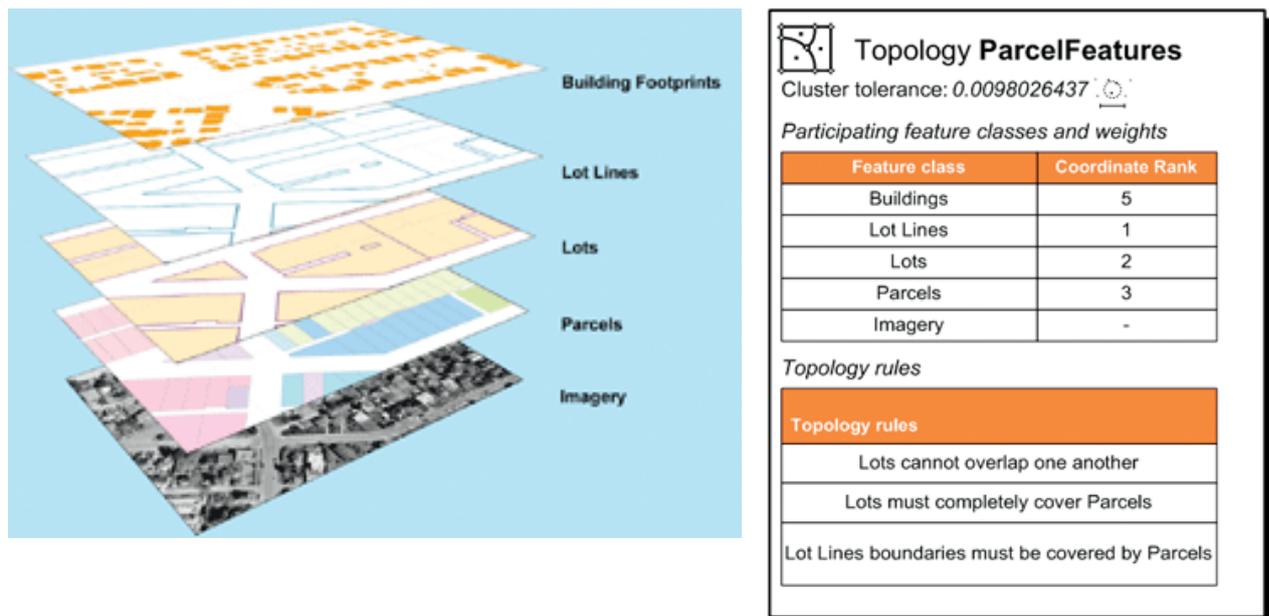


Figure 2.14. Topology rules applied (ESRI, 2002b)

According to Price (2010), whenever relationships between features are used frequently, time could be saved if explicit information about the relationships between those features is pre-defined. However, because a wide amount of topological relationships could exist between features in ArcGIS, the process of doing so are very flexible. The provision of a topology toolbar in ArcGIS provides tools that could be applied for the editing of features, preventing errors in the editing process and also allows the editor to find and correct previously made topology errors (ESRI, 2006).

The White Paper of ESRI (2005b) indicates a few important points regarding topological applications in GIS. These are:

- topology is applied in the management of shared geometry, in other words it enforces how features share geometry such as parcels that share edges;
- data and integrity rules are enforced for example when there need to be assured that no gaps would exist between features that connect, to prevent features to overlap or to ensure that endpoints connect if they need to be connected;
- topological relationship queries and navigation are supported which means that the ability to identify adjacent and connected features are enforced as well as to find shared edges, to navigate along a series of connected edges etc.;

- it provides editing tools that consist over advanced capabilities to enforce the abilities of the data model, for example in case a shared edge need to be edited, all the features that share the same edge will also be updated; and
- topology also consists over the capability to create features from unstructured geometry for example to arrange a few disorganized lines in order to construct a meaningful shape or polygon.

Applying topology

With reference to Table 2.6 an indication of the applications of topology in the geodatabase is provided. In the first column from the left, topology rules between features in the same feature classes are represented and in the three remaining columns, rules between more than one feature classes are shown. In every column the application of the topology rules are indicated. Among all the spatial relationships that could be applied, adjacency, coincidence and connectivity are validated with the creation of topology rules (Xie, 2006). When a topology rule between features is established, the editing of those features is limited to the rules created for them. In other words, topology rules provide a framework for features to be created. For example, if the rule “Endpoint must be covered by” is created for lines and points then it means that all endpoints should be covered by lines. Whenever a point is created without being covered by a line an error would occur and the operation would not be executed.

The three types of topology that are available in the geodatabase are:

- geodatabase topology – defines the spatial relationships for your data;
- map topology – during an edit session, temporary topological relationships among features in one or more feature class is created; and
- geometric network topology – stored in the same feature dataset, geometric network topology between point and line feature classes are created and stored.

Topology is created from feature classes stored in a feature dataset and each feature class can only participate in one topology at a time. Every feature class needs to be located in a feature dataset for a feature class to actively have topology rules defined for it (Xie, 2006).

 <p>Points</p>	 <p>Points on points</p>	 <p>Points on lines</p> <p><i>Must be covered by endpoint of Point must be covered by line</i></p>	 <p>Points on polygons</p> <p><i>Must be properly inside polygons Must be covered by boundary of</i></p>
 <p>Lines</p> <p><i>Must not have dangles Must not have pseudo-nodes Must not overlap Must not self overlap Must not intersect Must not self intersect Must not intersect or touch interior Must be single part</i></p>	 <p>Lines on points</p> <p><i>Endpoint must be covered by</i></p>	 <p>Lines on lines</p> <p><i>Must not overlap with Must be covered by feature class of</i></p>	 <p>Lines on polygons</p> <p><i>Must be covered by boundary of</i></p>
 <p>Polygons</p> <p><i>Must not overlap Must not have gaps</i></p>	 <p>Polygons on points</p> <p><i>Contains point</i></p>	 <p>Polygons on lines</p> <p><i>Boundary must be covered by</i></p>	 <p>Polygons on polygons</p> <p><i>Must be covered by feature class of Must be covered by Must not overlap with Must cover each other</i></p>

Table 2.6. A conceptual view of topology rules (ESRI, 2000)

The advantages of ArcGIS geodatabase topology

The motivation of data consistency and error free editing of data within the boundaries defined by topology rules motivates the application of topology within the geodatabase. The two worlds of editing and data deployment were united in the geodatabase by means of topology management. The advantages are:

- topological queries;
- better data management;
- shared geometry editing;
- rich data modelling;
- improved data integrity;
- more flexibility;
- a simple, highly scalable data storage mechanism based on open, simple feature geometry; and
- a fast, simple, and efficient data model which can be edited and maintained by many simultaneous users.

(ESRI, 2005b)

Theobald (2001), stated, “one of the main reasons that promoted the development of topology was to provide a rigorous, automated method to clean up data entry errors and verify data.” The data entry errors he mentioned are errors prone to slivers, dangles, and over- and under shoots. Topology would therefore serve as a medium through which these errors could be corrected by means of rebuilding the topology. It also serves as a data inspection system that could be repeated as many times as necessary.

Topological data structures are much smaller than compact file sizes due to the fact that shared vertices or boundaries of neighbouring polygons are not stored twice. This means that the limiting storage factor is not a problem anymore. Also, the process of finding adjacent features is simplified with the presence of topology stored in the system.

The conclusion of the aforementioned section finally supports the motivation that topology are applied to model real-world features in a better way (Price, 2010). It defines and enforces data integrity rules to ensure that the input of the data is valid according to the predefined rules and standards of the geodatabase. Furthermore, topological relationship queries and navigation are supported by topology and it provides sophisticated editing tools which ensure the creation of structured features from unstructured geometry. Ross & Cleveland (2005), motivates that topology provides the ability for the geodatabase to model geometric relationships in the real world more accurately. Topologies are simple and easy to use and provide simultaneous editing of features.

Topology in 3D GIS

Topology is an extremely powerful function that is applied in GIS especially in a 2D environment. However, topology inside a 3D environment is approached in a different way. Since 1999, the use of 3D in GIS became more significant. Various kinds of products delivered fascinating results and 3D representations, however, their application within a GIS environment was limited. With the release of ArcGIS 9 and the 3D applications it provided, the 3D GIS environment was enhanced more than any other software package previously released (Smith & Friedman, 2004).

Shortly after the release of ArcGIS 3D, data in OpenFlight, 3D Studio and VRML formats could be imported. Within a few weeks from the release of ArcGIS 9 different plug-ins for various software programmes were created in order to import drawings directly in GIS (Smith & Friedman, 2004). With the development of 3D GIS, users have the ability to present

ideas and designs more realistically. The interior and exterior view of buildings could be modelled. Each floor can be modelled as a layer with separate attributes relating to that floor only. With 3D GIS (Virtual GIS), the user is provided with the ability to visualize real world phenomena in detailed 3D views. It also enables the user to move to different locations in an area. The layout of buildings and streets, as well as utilities, could be viewed on their actual topography and analyzed in virtual 3D GIS (Koller et al. 1995).

It can be said that the functionality of 3D GIS provides much more benefits than that of the 2D GIS representations when referring to the realistic representations of real world phenomena. However, 3D GIS have not yet reached the point where its functionalities have been as well developed in all areas than that of the 2D GIS world. A major difference in the two systems is the fact that topology and the implementation of topology rules in the 2D GIS world are well developed and applied (Ellul & Haklay, 2006).

Unfortunately, the same functionalities have not yet been as well developed in 3D GIS. Up till now the focus of the 3D GIS has mainly been on the visualization functionality it provides and therefore, the topological functionality of 3D GIS has not yet been well developed. Lee (2004) stated that even though commercial GIS extensions and products can handle 3D data with regard to visualization, it still lacks high-quality functionality in the areas of 3D spatial data structuring, data manipulation and data analysis. According to Lee (2004), 3D GIS have originally been developed with the focus on integrating 2D GIS functionalities with 3D CAD geometric representations. Lee (2004) discovered two problems with 3D GIS data models. Firstly, due to complex geometric computational problems these 3D data models lack efficiency in maintaining topological consistencies. Secondly, there are also problems with the connectivity relationships between boundary representations which are not clearly stored. This problem prevents efficient network-based analyses and also impede on sufficient route analyses in 3D geographic entities. According to Yuan & Zizhang (2008), execution of indoor navigation in 3D could not yet be performed as expected due to the fact that the topological structure for the indoor environment are not yet well developed.

Shephard (2009) stated that topologies are well developed and maintained in a 2D GIS environment. Unfortunately the 3D GIS environment still has a lot of scope for development in that area according to Kuehne (2010). He also stated that there will not be any topology available in ArcGIS 10.1 either. According to a forum post by Murphy (2010), ArcGIS are

only capable to detect topological errors in 2D GIS environments and not yet in 3D GIS environments.

2.4.1.5. Geometric networks

In connection with the previous elements – datasets, relationships, subtypes, domains, and topology – two very important aspects that all of these elements have in common are connectivity and relationships. Connectivity between features implies that some sort of connected relationship between those features exists. Nevertheless, these two aspects serve as the key features when referring to networks, providing an organized, understandable and structured way of representing them (ESRI, 2006).

According to the Environmental Systems Research Institute (ESRI, 2006), there are two main categories in which networks could be classified: physical networks and logical-social networks. Price (2010) oppositely implied another category for networks, which combines the relationship between these two categories as a representation in ArcGIS. This category is geometric networks in combination with an attribute based logical network. The combinations are indicated in Figure 2.15.

These examples are modelled in ArcGIS as one-dimensional non-planar graph or geometric network that is composed of “features” which is why these features are considered to be network features (ESRI, 2005d). In ArcGIS, topological relationships in these geometrical networks are automatically maintained by means of a logical network (Price, 2010). An important characteristic of geometric networks that need to be stressed is that it contains a logical network. This network could be defined as a “cloned geometric network” due to the fact that it is created at the time a geometric network is created and resembles the geometric network with the focus on providing advanced capabilities to the geometric network. This means that a logical network represents and models the connectivity between features and serves as the connectivity graph used for tracing and flow calculations. Furthermore, it maintains connectivity between all the edges (lines) and junctions (points) in a feature dataset (ESRI, 2006). The logical network, as stated by Price (2010), contains information regarding the construction and operation of the network elements. In more basic terms, it stores the documented information about the relationships, connections and the behaviour of the geometric network in tabular format. This tabular data provides an indication of how features involved in the geometric network are connected to each other (ESRI, 2006).

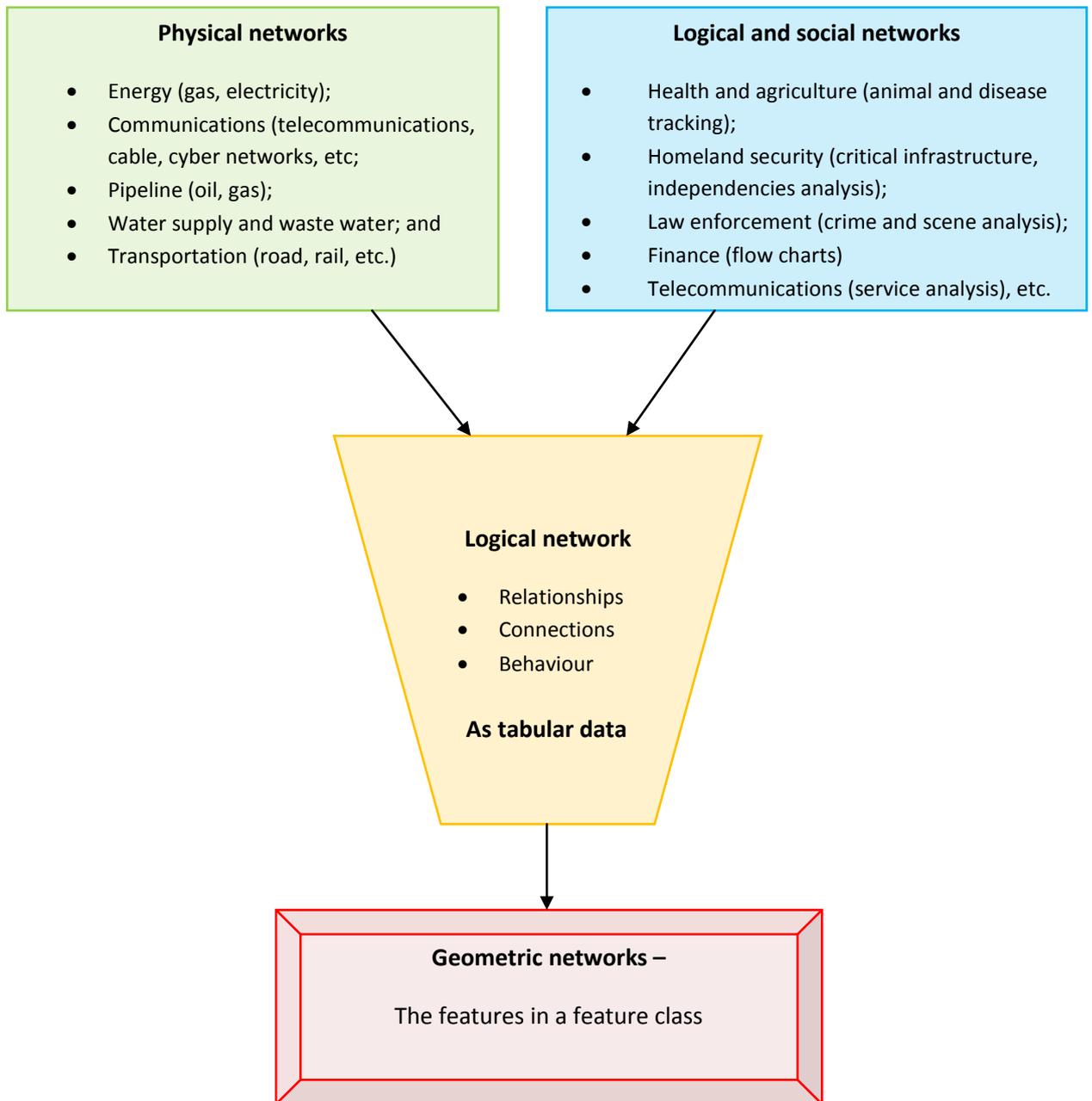


Figure 2.15. The connection between different networks (Price 2010; ESRI, 2005d)

The logical network has limited access for the user because it is created in the time the geometrical network is created, built from the feature classes. The logical network and the geometrical network are therefore interconnected, which means that changes in the geometrical network require adjustments in the logical network but these changes are maintained automatically. During the processes of editing and analysis, the logical network provides quick realization and modelling of the connectivity between the edges and junctions of features in a geometrical network. This capability ensures fast tracking and network

connectivity and provides the geometric network with a capability like no other. Advanced editing options with regard to geometric networks are also available. These include

- directional indications for the flow of resources through a network; and
- weights could be assigned to different features in order to control the speed at which commodities flow through different parts of the network.

As mentioned by Price (2010), many questions with regard to network systems and the service they provide, could be answered through modelling the behaviour of various network features with geometric networks. An example scenario would be to indicate which area in the town would be affected if a broken segment in the pipeline occurs. Locating the damaged segment, knowing which valve should be closed down that would affect the smallest amount of people in the town, are information that could be retrieved with the help of a geometric network. Modelling network systems with geometrical networks could have a lot of advantages in the areas of maintenance and management. Similar cases are summarized with the analysis versus the application in Table 2.7.

According to a product manager and user advisor for ESRI ArcGIS, Law (2009) defines a geometric network as “a set of connected edges and junctions, along with connectivity rules that are used to represent and model the behaviour of a common network infrastructure in the real world.”

Geometric networks are defined by feature classes inside the geodatabase that serve as the data sources. Inside the geodatabase the features act as units for the geometric network that act upon the roles assigned to each one of them. There are also rules that specify how the resources should flow through the geometric network. The flow of resources through the geometric network has specified rules assigned to them that ensures the correct flow in the network.

Analysis	Application
Calculate the shortest path between two points.	Various kinds of utility companies use this as a method of inspecting the logical consistency of a network and verifying connectivity between two points.
Find all connected or disconnected network elements.	Electric companies can see which part of the network is disconnected and use that information to figure out how to reconnect it.
Find loops or circuits in the network.	An electrical short circuit can be discovered.
Determine flow direction of edges when sources or sinks are set.	Managers or engineers can see the direction of flow along edges, and ArcGIS can use the flow directions to perform flow-specific network analyses.
Trace network elements upstream or downstream from a point.	Water utilities can determine which valves to shut off when a pipe bursts.
Calculate the shortest path upstream from one point to another.	Environmental monitoring stations can hone in on a source of pollution in streams.
Find all network elements upstream from many points and determine which elements are common to them all.	Electric utility companies can use the phone calls of customers experiencing an outage to locate suspected transformers or downed lines.

Table 2.7. An analysis of geometric networks versus real life applications (ESRI, 2006)

As previously stated, geometric networks are represented by points and lines in ArcGIS. The most basic representation of this concept for geometrical networks is indicated in Figure 2.16. The lines in Figure 2.16 represents the edge network features and the dots are examples of the junction network features. The edges need to be connected to other edges through junctions. A summarized version of the edges versus junctions is represented in Table 2.8. Although topology and geometric networks are total different data structures edges and junctions are

topologically connected to each other. Therefore edges need to connect to each other at junctions.

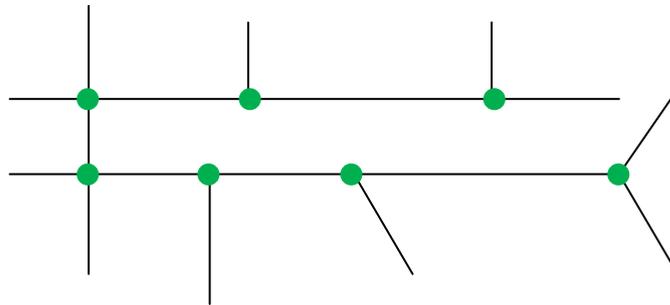


Figure 2.16. A geometric network in its basic form (Self created)

A summary of the edges versus junctions	
EDGES 	JUNCTIONS 
<ul style="list-style-type: none"> ▪ Provides length through which a resource flows ▪ Created from line feature classes ▪ Correspond to edge elements in a logical network ▪ Examples include, water mains, electrical transmission lines, gas pipelines, telephone lines 	<ul style="list-style-type: none"> ▪ Serves as a connection for two or more edges to connect ▪ Enables the transfer of flow between edges ▪ Created from point feature classes ▪ Correspond to junction elements ▪ Examples include, fuses, switches, service taps, valves

Table 2.8. A summary of the edges versus junctions (Law, 2009)

Both edges and junctions have sub-features called simple edges and complex edges, user defined junctions and orphan junctions as indicated in Tables 2.9 and 2.10.

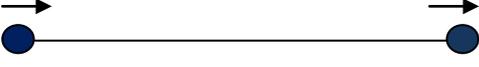
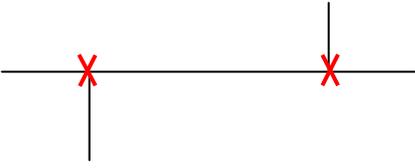
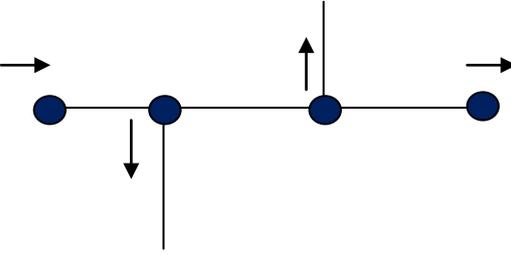
Simple and Complex edges	
Simple Edges	Complex Edges
<p>Enters Exit</p>  <p>Midsnap snapping or any snapping along the edge would result in an error or the edge would be divided in two or more sections.</p> 	 <p>Complex edges allows midsnap snapping and the feature would remain a single feature after midsnap snapping.</p>

Table 2.9. Simple and complex edges (Self created)

User-defined and orphan junctions	
User-defined junctions	Orphan junctions
<p>These junctions are created based on a user's source data (point feature classes)</p> <ul style="list-style-type: none"> • Service points • Fuses • Stream gauges • Taps 	<p>At the creation of a geometric network a corresponding simple junction feature class is created called the orphan junction feature class. The name of the orphan junction relates to the name of the geometric network added with a _Junctions suffix. Eg. Network dataset = Electric_Net Orphan Junction Feature class = Electric_Net_Junctions</p> <p>Orphan junctions are inserted at the endpoint of every edge at which a geometrically coincident junction does not already exist.</p>

Table 2.10. User defined and orphan junctions (Self created)

As mentioned by Xie (2006), Figure 2.17 represents a layout of the interactive layers in a geometric network. Each feature represents either a point or a line. The end result is composed of all the feature classes participating in the geometric network

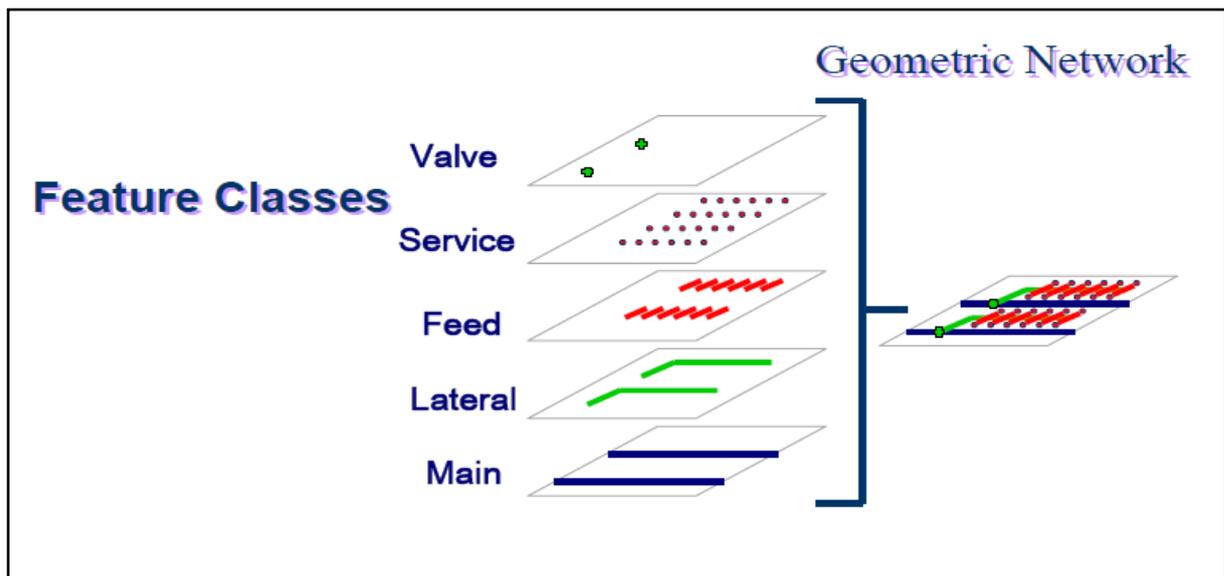


Figure 2.17. A layout of the geometric network (Xie, 2006)

Geometric networks are generally applied in the utility environment that includes electrical, gas, sewer and water lines and allows travelling along edges in one direction at a time. An example of the former mentioned is the direction in which oil in pipeline flows, which is primarily determined by the external forces of gravity, electromagnetism, water pressure, and so on. These elements could be controlled by an engineer who could control how these external forces act on an agent flowing through the pipeline (Cummins, 2011).

Geometric networks model utility networks, which differ from transportation networks in two cases. Firstly, a utility network is a directed network, in other words there is a fixed direction for the commodities that flows through the utility network, managed by a set of predefined rules. The second difference between the two network models is the fact that the geometric network allows automatic updates of the utility network whenever a source feature in a network is changed. However, both the network dataset and the geometric network have a slight disadvantage due to the fact that it could become time consuming to rebuild the network whenever a change is applied to the source, especially in large geometric networks (Veenstpe, 2011). According to Maren (2011), 3D geometric networks are not supported in ArcGIS 10, however, the shortest route between two points in 3D could be found by working in ArcScene. With reference to Table 2.11 a summarized version of the network datasets versus the geometric networks are provided.

Network datasets vs. Geometric networks	
Network dataset	Geometric network
For transportation modelling	For utilities and natural resources modelling
Path finding and allocation operations	Network tracing functionality
Turns supported	Turns not supported
Uses simple feature: points and lines	Uses custom features: simple/complex edge features and junctions
A more robust attribute (weight) model	Weights based only on features fields
User controls when connectivity is built	System managed connectivity

Table 2.11. A summary of the network datasets versus the geometric networks (Veenstpe, 2011)

Both geometric networks and network datasets have been considered for the current study, however, it has been found that a geometric network cannot be represented inside a 3D environment. For this reason and the fact that the features in a network dataset could support topology, the network dataset have been chosen to represent utilities in this study.

2.4.1.6. Network datasets

Connectivity between features is not just established by means of a geometric network but also by a network, dataset within a geodatabase. When referring to transportation networks that include streets, pedestrian and railroad networks, movement along the roads or railways (edges) could occur in two opposite directions (Cummins, 2011). Network datasets are also created from source features, including simple features such as lines, points, and turns. Connectivity of the source features is stored inside the simple features and the turn features in a network dataset. As stated by ESRI (2010c), any analysis done in ArcGIS Network Analyst, are always executed in a network dataset.

The Environmental Systems Research Institute (ESRI, 2010c) explains that connectivity between features is compulsory because the features in a system are usually unaware of each other and therefore needs some kind of specification. The network dataset enables the features to keep track of each other by storing properties of the features and their behaviour with reference to each other. The network dataset has a connectivity policy that one can

modify in order to define which coincident features are truly connected. Overpasses and underpasses for streets could therefore also be modelled.

The network analyst extension in ArcMap relies on the network dataset as the primary data model. The features in the network dataset do not have any specialized behaviour and therefore topology could be defined for it which is not the case for a geometric network. Feature classes could participate in multiple network datasets but a feature class could not participate in both a network dataset and a geometric network (Veenstpe, 2011). The network dataset that models transportation networks models undirected networks, in contrast with geometric networks that models directed networks.

With ArcGIS a persistent network is stored, providing the possibility to modify its properties and model a variety of networks using network datasets. According to ESRI (2010c), the best option to create a network dataset is from the feature classes in a feature dataset of a geodatabase. The feature dataset consists over the ability to easily communicate with multiple feature classes, supporting multiple sources and model a multimodal network.

2.5. CAD versus GIS

With reference to facility management, water utilities, engineering and construction disciplines; different software programmes could be applied to solve problems in each of these areas of expertise. Every software application could provide a different and unique way of interpreting the various disciplines, with a different focus for each. The integration of several systems, each addressing a different part of a problem, would result in a holistic and effective solution. Some software systems do have the same approach and might even deliver a result in the same format, such as CAD and Microstation. However, finding a way of integrating two different systems in order to deliver a complete solution would be the ultimate approach. It is about analyzing the problem, approaching it from different perspectives, addressing different areas of the problem, finding the best solution in the most effective way. This also means that the programmes applied should not necessarily be homogeneous in all of their ways. Problems with heterogeneous environments should be approached from different views and therefore interoperability between different programmes needs to be established in order to apply interoperable tools and methodologies. Peachavanish et al. (2006) however stated that this could become a very complex situation.

For the purpose of this study, concerning the integration of spatial and non-spatial data and combining data of utilities and geographic location, emphasis would be placed on two heterogeneous programmes, Computer-Aided Design (CAD) and Geospatial Information Systems (GIS). Although CAD and GIS have many areas where they might correspond in the structure and the way they manage data, these programmes have been developed independently over the years with different purposes without any means of operating together. According to Akinci et al. (2008), significant differences between CAD and GIS exist in the data formats each programme supports, terminology each one utilizes, semantics of concepts they represent and the reasoning techniques on which they are based. Therefore, integration between these software systems is not a simple procedure, but do complement each other by means of data sharing and data conversion whenever correctly applied.

2.5.1. CAD

Computer-Aided Design (CAD) drawings made its appearance on personal computers in the 1980s, 20 years after the basics of the system were developed in the 1960s. Within ten years the technology became widely known and established, and was applied in everyday use in the departments of architecture and engineering practices. Complex information of larger drawings and CAD data in these departments had to be managed and this was done by means of organizing the data into different layers. This way graphical drawing elements of the same type could be allocated into “invisible” layers and provides the user with the ability to switch between different layers or even turning selectable layers on and off. This process enables the user to keep focus on the features essential to the work at that stage (Figure 2.18) (Howard & Björk, 2007). According to Morgan (2009) CAD drawings are the conversion of the classical transparent map into an electronic version of objects, grouped together into a set of layers.

According Maguire (2003), Computer-Aided Design (CAD) focuses on automating the drafting process with electronic drawings. It is a medium for representing the design of an existing object, primarily in a 2-dimensional digital format, with the focus on 2D graphics, sketching and co-ordinate geometry tools. However, recent software technology includes 3-dimensional data modelling capabilities and object-oriented design (Morgan, 2009) for modelling man-made things (Pu & Zlatanova, 2006). CAD systems are very versatile, providing functions to update, edit and display features in different ways (Cowen, 1988). Typical applications of CAD data include project engineering and site design, facilities management and construction drafting (Maguire, 2003). These CAD drawings store and

exchange vector-based data and contains no additional attributes. As mentioned by Cowen (1988), the CAD system cannot automatically execute or perform actions on entities based on values stored in a database such as with GIS, because it is hard to link attributes in a database to specific geographical objects. Maguire (2003) mentioned that the graphic in a CAD drawing is the database and a drawing file is the persistence container in which the graphics are stored.

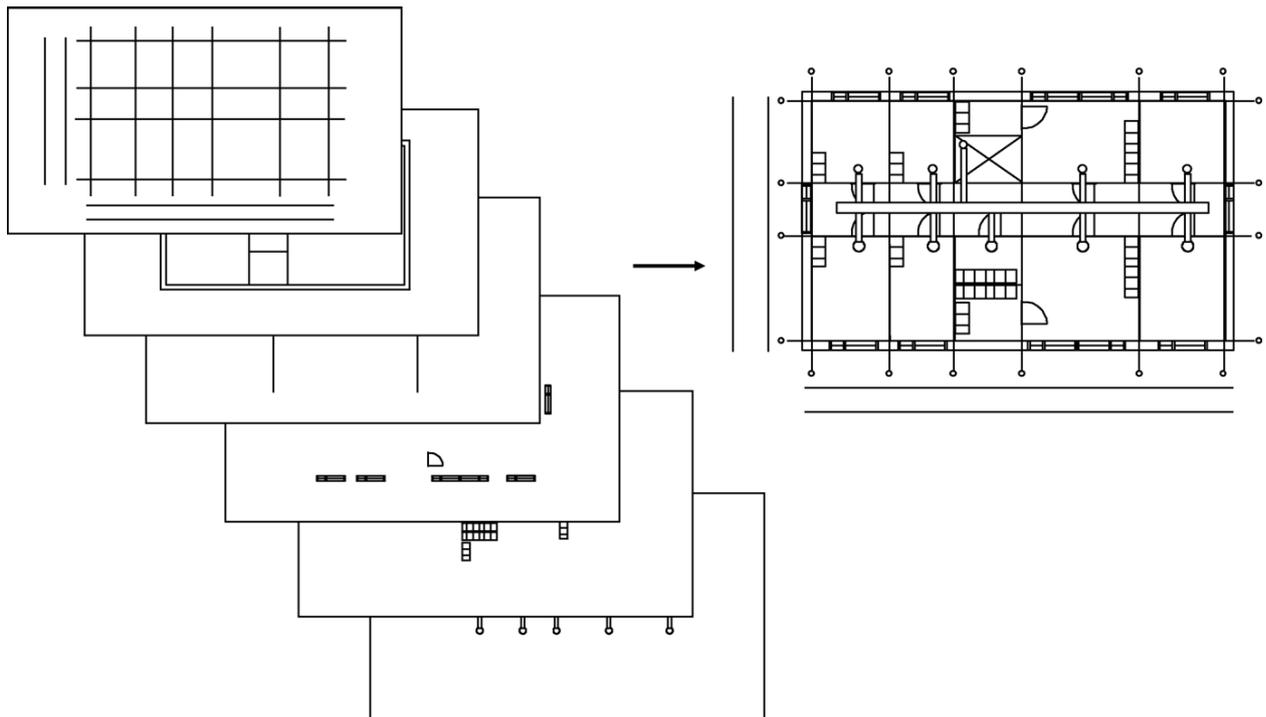


Figure 2.18. Representing CAD data as a number of transparent sheets (Howard & Björk (2007))

Very accurate drawings and operations could be executed with a CAD system, but it contains no database management system to promote automatic executions according to attributes inside a database. CAD drawings are also file based documents and are not stored in a database with advanced editing functionalities such as with GIS. Even though real world objects are represented by points, lines or polylines, the CAD system does not contain the functionality GIS provides. This explains why CAD is merely a graphic drawing system. However, CAD software and CAD drawings are generally applied in building and construction environments, such as engineering, architecture, town-planning and surveying. CAD drawings serve as documents such as legal proposed construction plans that define real

world objects. These plans could be used in accordance with other technological software systems such as GIS, where they serve as data collection documents.

For CAD-GIS interoperability, the spatial reference of the datasets needs to be addressed. GIS is known for its spatial location functionality, but CAD drawings are graphical drawings based on graphical representations with inconsistent co-ordinate systems. Overall, CAD drawings are based on random Cartesian co-ordinates relative to an object on the drawing such as a building. This implies that CAD orientated drawings cannot be represented on a map with real world co-ordinate systems and need to be georeferenced (Morgan, 2009). With regard to the aforementioned, there are some similarities between CAD and GIS. CAD and GIS correspond in areas of geometry although they differ in certain aspects with regard to size, storage, analysis, semantics, attributes, etc. Information and representations about real world objects are both provided by CAD and GIS.

2.5.2. The differences between CAD and GIS

Building and construction drawings are created for certain purposes and therefore finer detail in a drawing is not available for use in different software programmes, which implies that spontaneous interoperability are complicated (Morgan, 2009). In other words, the drawing in a CAD format serves as the actual information whereas with GIS, “the lines are just the representation of the data behind it” (Guerrero, 2007).

CAD drawings are graphical drawings with much less functionality than the drawings inside a GIS system and therefore the information required for a CAD drawing is basic and straightforward. In opposition, additional functionalities are presented in GIS. More formal data models with rule-based editing and topology are provided in GIS, with the emphasis on spatial analysis and high-end cartography Maguire (2003). The features inside of a GIS could contain vast amounts of data describing the features. For example, a polygon representing a shopping centre could contain information about the name and classification of the business, the owner or managers of the centre, the street address, and the number of parking facilities. But the most important difference between the two systems is that GIS data are spatially informed with regard to locations, adjacency and other spatial relationships. In basic terms, GIS is a spatial database operating system and CAD is a graphical drawing system (Berthiaume et al. 2005 & Guerrero, 2007).

As described by Pu & Zlatanova (2006), the original design of CAD focused on 3D tools for modelling human created objects (cars, industrial machine parts, buildings etc.), which were represented and designed in local co-ordinate systems. Ultimately, the focus of the CAD software was on its design concepts with emphasis on editing tools and effective 3D visualization. In contrast with the latter, the design of GIS systems focused on representing real world phenomena with main emphasis on all the tasks initially performed on paper maps (Zlatanova & Stoter 2006). The ability of GIS, therefore, supports the maintenance of points, lines, and polygons with geographic co-ordinates and corresponding attributes and is able to provide specific spatial analyses.

2.5.3. Problems of integrating CAD with GIS

Research done by Pu & Zlatanova (2006), Zlatanova, Rahman & Pilouk, (2002), and Zlatanova & Stoter (2006), found that there are still a few critical issues regarding the merging between CAD and GIS. Among these issues are:

- “CAD software supports a broad range of technological development such as cone, sphere, cylinder and free-form curves while these technological developments are not present in the GIS world” (Pu & Zlatanova, 2006);
- GIS packages contain limited 3D editing tools (Zlatanova, Rahman & Pilouk, 2002); and
- in both CAD and GIS systems there are a lack of 3D topology and therefore also in 3D analysis (Zlatanova & Stoter, 2006).

Since the original design of CAD and GIS, the focus of the programmes was different. Based on the findings of Van Oosterom (2004):

- *Representation:* Different objects about the man-made world are represented by CAD whereas GIS specialized in representing the natural environment. This implies that the mathematical descriptions of the two systems are quite different. CAD maintains the ability to represent complex detailed drawings very accurately where the focus for GIS is on capturing, storing and analyzing large amounts of data inside a database.
- *The timescale:* CAD systems operate on a project basis, the lifecycle maintenance of the data is fairly short and therefore the data is stored in a file format. However, data collection and maintenance in GIS, spans over a much longer period with almost an

endless life cycle. The data of GIS are consistently and permanently maintained inside a geodatabase.

- *The represented focus:* CAD systems represents 2D and 3D drawings based on man-made structures while GIS systems deal with data sources based on many different co-ordinate systems, which are used to model the spherical (ellipsoid or geoid) world (Van Oosterom, 2004).

One major common characteristic between the two systems is the fact that they both deal with geometry. However, there are still many differences in the way both programmes are created, the organizations that maintain it and the target purposes thereof. This in turn implies differences between the two systems in the areas of their size, storage, analysis, semantics and attributes (Van Oosterom, 2004).

With the development of technology, integration systems have been designed to integrate CAD and GIS mediums. In the article of Van Oosterom (2004), a solution for integrating CAD/GIS was identified. This included a way for solving semantic differences between the two systems and to create an integrated model that would be able to maintain consistency during updates or the model could be added to the database management system (DBMS). In basic terms, the model is created to serve as a mutual source where data is available to all users.

2.5.4. Choosing CAD as main data source

GIS is in use for some time now and are growing in its popularity as a tool for various disciplines. However, the lack in accurate GIS data representation of different features for certain areas is a reality. Since the 1980s, manual drafting have been replaced by Computer-Aided Design (Bureau of Labor Statistics, 2008). CAD has therefore been applied in different guises throughout various departments such as architecture, engineering and manufacturing. These applications led to collections of project-oriented construction drawings from previous projects that are available in CAD format (Berthiaume et al, 2005). At a larger scale these CAD drawings are also very accurate, based on a ground survey undertaken by a certified professional surveyor. Detailed information about facilities are captured and represented using CAD drawings after which it could then be imported into GIS in order to perform location related analysis of those components at different scales (Peachavanish et al, 2006).

With regard to the corresponding characteristics between CAD and GIS, certain facts have already been stressed in order to motivate the interoperability between the two software systems. Van Oosterom (2004) mentioned a few points which empower the importance of integrating CAD with GIS:

- *Plan development:* With the design of large infrastructures both CAD and GIS are needed. While CAD handles the engineering and construction techniques GIS would cover the data that is essential for planning and layout of the structures. These two systems are constantly interdependent of each other. Geographic data are imported into CAD for design and vice versa.
- *Visualization:* Whenever data are prepared for a project it is important to visualize different views of the data required for plan representation and data interaction. These representations include: 2D “plan view” for initial context analysis, a 2.5D “model view” and a 3D “world view” for more realistic visualization of the following design.
- *Data collection:* The process of data collection has evolved through the years from manual editing to remote sensing and photogrammetry. Some photogrammetric techniques are well managed in CAD because it recognizes different landmarks and features in a landscape. Features that cannot be surveyed from the outside such as underground utilities appeal to the use of CAD drawings in 3D GIS modelling.
- *Location based service:* A combination of CAD and GIS techniques need to be combined for accurate positioning of locations and the provision of directions and appropriate sight information. However, there is still ample space for improvement between the two systems before automatic notifications of an integrated CAD/GIS model would be presented in the way a GPS represents located information.

2.6. Georeferencing CAD data

The process of scaling, rotating, translating and deskewing an image to match a particular size and position is called georeferencing. “It is also the process of defining how raster and vector data is situated in map co-ordinates” (University of Alberta, 2010) by “assigning values of latitude and longitude to features on a map” (Alumbaugh & Bajcsy, 2002). According to Kurnia, Hopkins & Ho (2005), the features in a GIS must be georeferenced so that data can be correctly related to locations on the earth’s surface, in other words, referencing a map image to a geographic location, transforming 3D co-ordinates to and from 2D map co-ordinates.

There are mainly three possible classes of transformations as described by (Alumbaugh & Bajcsy, 2002). These are:

- 2D-to-2D transformations, the process of aligning points that represent identical real world locations.
- 3D-to-3D transformations, the process of integrating information such as the combination between two or more different datasets.
- 2D-to-3D transformations, the process of identifying the location of a feature on 2D digital terrain map if it is identified in a 3D latitude/longitude co-ordinate environment.

The importance of georeferencing a drawing is to align the raster or the vector dataset to a map co-ordinate system for viewing, querying and analysis with other data layers (University of Alberta, 2010). On a georeferenced map, it is important to be able to retrieve the latitude and longitude co-ordinates for any *point* on the map.

For the current study, it was important to apply georeferencing to the CAD data in order to ensure accuracy when images and CAD data had to be aligned. It also ensures that data are orientated according to real world locations and precise positions could be located. If a CAD drawing is georeferenced it also enables the user to digitize information from the drawing from where the drawing could be used as a visual display in a map (Parmenter, 2005).

ArcGIS provides a special toolbar to move, rotate and scale CAD layers with simplicity. This means that the source and destination co-ordinates of the transformation could interactively be specified by using control points (ESRI, 2007d). Whenever a CAD drawing needs to be georeferenced according to a raster backdrop image, it is important to establish reference points (Navya, 2011). This means that there need to be two points on the CAD drawing that correspond to the same points on the raster image. ArcGIS contains a georeferencing toolbar with georeferencing tools which enables the user to execute georeferencing procedures with good accuracy.

2.7. Co-ordinate systems

Geographic co-ordinate systems serve as a framework for defining real world locations and to assign geographic locations to objects. Latitude-longitudes serve as a global co-ordinate framework of which a planar Cartesian co-ordinate system is derived from. Locations on the earth's surface are measured with latitude and longitude lines in a projected co-ordinate system. "GIS datasets contain co-ordinate locations within a global or Cartesian co-ordinate system to record geographic locations and shapes" (ESRI, 2009).

In this section geographical co-ordinate systems (GCS) and projected co-ordinate systems will be discussed as well as the associated components thereof.

2.7.1. Geographical co-ordinate systems (GCS)

"Co-ordinate systems serve as a medium to locate geographic features on a two- or three-dimensional surface" (Kurnia, Hopkins & Ho 2005). With the help of a geographical co-ordinate system (GCS), the problem of disorientation has been solved quite some time ago. "With GCS a three dimensional surface are applied to define locations on the earth" (Kennedy & Kopp, 2000). Breaking the concept, geographical co-ordinate systems down into smaller parts, the components are as follows: an angular unit of measure, a prime meridian and a datum that is based on a spheroid. Looking further into these concepts, there are another few points of focus, latitude and longitude. These concepts describe the angles that are measured from the earth's centre to a point on the earth's surface often measured in degrees.

The latitude lines in the spherical system are known as the East-West lines or parallels and the longitude or meridians are known as the vertical lines or North-South lines and together these lines cover the earth and form a network known as a graticule (Figure 2.19). The equator is known as the line midway between the poles, zero latitude and the line of zero longitude is known as the prime meridian. Where the equator and the prime meridian intersect, the graticule originates (0, 0). Furthermore, the earth is divided into four quadrants, North, South, East, and West respectively.

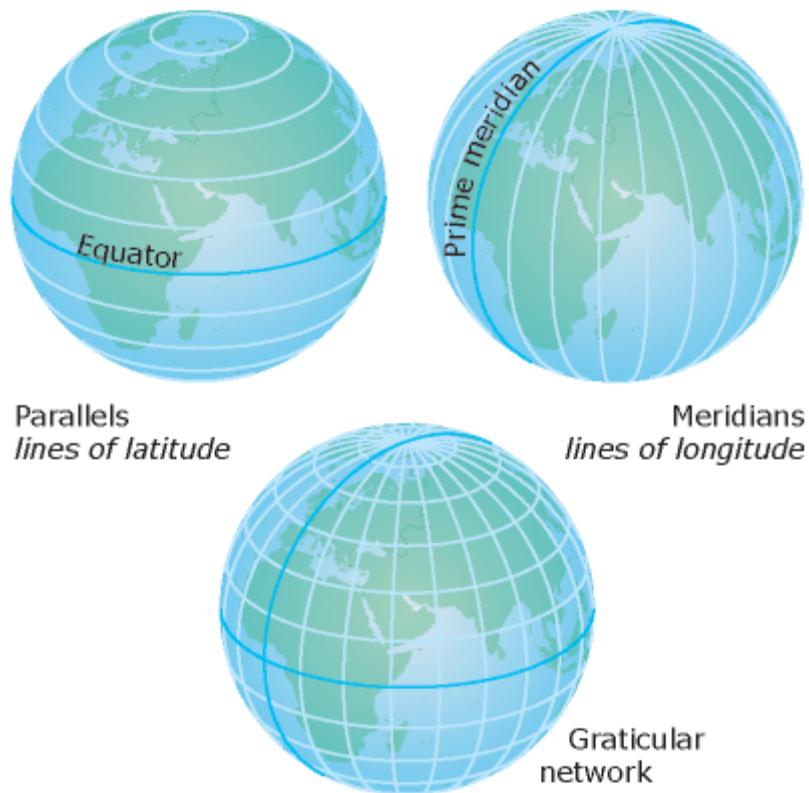


Figure 2.19. A representation of latitude and longitude lines (ESRI, 2009)

Because the shape of the earth is more clearly described as a sphere rather than a round circle, the distances in longitude and latitude also differs. Relative to the equator, the latitude vales ranges from -90° at the South Pole to $+90^{\circ}$ at the North Pole. Longitude values on the other hand, are measured relative to the prime meridian with values ranging from -180° west to $+180^{\circ}$ east. Unfortunately, values are only exact along the equator and become less accurate when moving away from the equator (Alumbaugh & Bajcsy, 2002).

2.7.2. Spheroids, spheres and datums

When looking at the shape of the earth it is resembled through a spheroid and serves as a mathematical model for the earth. During the years many interpretations of the earth have been established with the use of satellite technology. Representations of the earth have been produced but there are some influencing factors, such as the gravitational and surface feature variations, which imply that the earth is not a perfect sphere or spheroid. This could have an influence on the accuracy of locations on the earth. The reason is that “a local datum aligns its spheroid so closely to a particular area on the earth’s surface; it is not suitable for use

outside the area for which it was designed” (Kennedy & Kopp, 2000). Any position on the spheroid relative to the centre of the earth is defined by a datum which serves as a frame of reference for measuring locations on the surface of the earth. “A datum also defines the origin and orientation of latitude and longitude lines” (Kennedy & Kopp, 2000).

As indicated in Figure 2.20 the geoid resembles the earth without any topography and the ellipsoid (also known as the spheroid) resembles the earth as a simplified version of the geoid. In Figure 2.20 the different components are placed over each other.

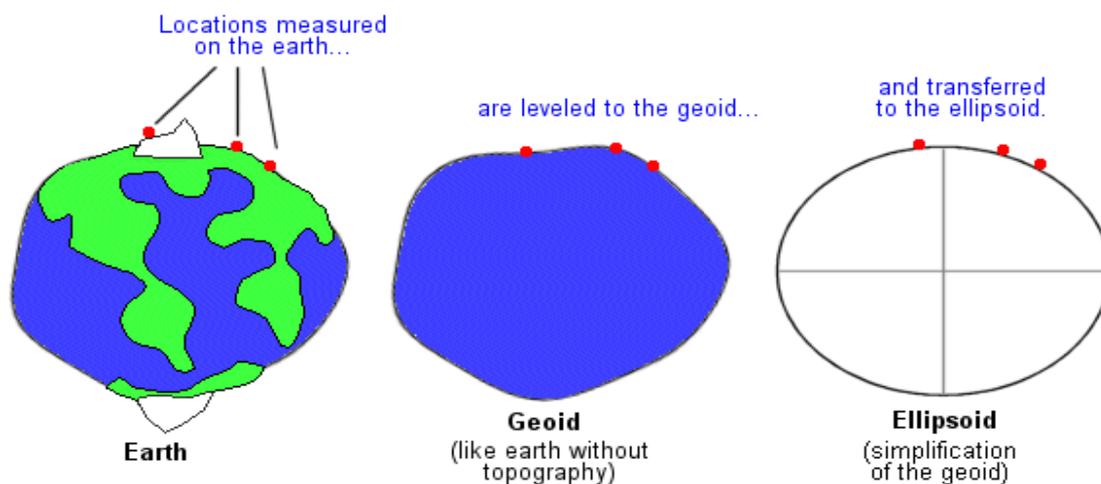


Figure 2.20. A representation of a geoid and an ellipsoid (Steiner, 2007)

With reference to Figure 2.21 whenever the geoid and the spheroid match a particular position the origin point of a datum is created. This point serves as the fixed point from which all other points are calculated from. However, two datums could be created; earth-centred datum and a local datum. The difference between the two is that the earth-centred datum (also known as a geocentric datum) is created whenever the spheroid aligns with a point on the middle of the earth. In other words, the earth’s centre of mass serves as the origin of the earth-centred datum. The WGS 1984 datum is currently one of the general used datums and is used worldwide for local measurements. From a theoretical perspective, it has also been found that this datum is the one most applicable for the current study.

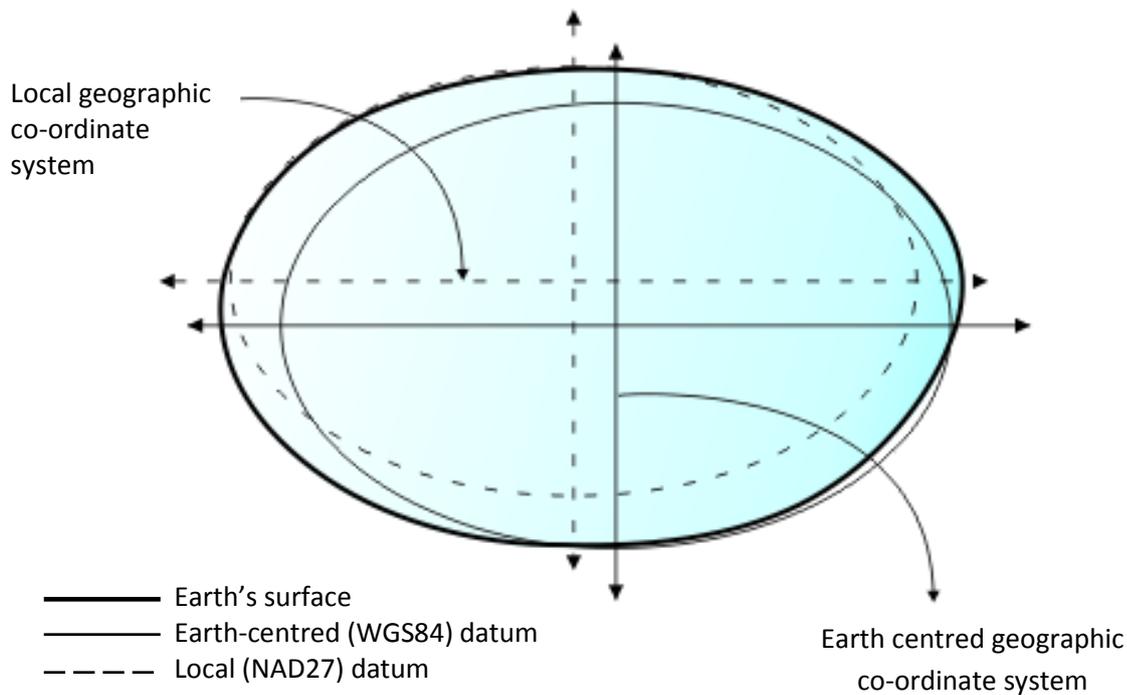


Figure 2.21. The geoid (IBM, 2010)

Local datums on the other hand are created whenever the geoid aligns with a point on a particular area on the spheroid. This point is not earthly centred as in the case of the geocentric datum and are more applicable to areas it is designed for. For this reason inaccuracy of co-ordinate system would occur if a datum is used for an area it was not originally designed for (ERSI, 2008b).

In the South African context the Cape datum, Hartebeeshoek94 datum and the WGS 1984 datum are all relevant datums that could be applied for the current study. On the 1st January 1999, the general South African co-ordinate system that was used changed from the Cape Datum that were based on the Clarke 1880 ellipsoid with its origin points at Buffelsfontein near Port Elizabeth to the Hartebeeshoek94 Datum based on the WGS84 ellipsoid with the IRTF91 (epoch 1994.0) co-ordinates of the Hartebeeshoek Radio Astronomy Telescope as origin. The reasons for the transformation were the advancement in technology, especially in the area of modern positioning (GPS) technology which made the flaws and distortions in the current system apparent (TechnoCAD South Africa, 2011). The transformation brought along many advantages that were consistent and near distortion-free. The new system is also well placed in a global scene, which makes it virtually compatible with the GPS reference frame. The Hartebeeshoek94 co-ordinate system serves as a uniform South African network

including areas such as Namibia, Botswana, Zimbabwe, Mozambique, Lesotho and Swaziland (Wonnacott, 2000).

2.7.3. Projected co-ordinate systems and projections

Map projections are one of the core aspects when referring to GIS and geodatabase design. Everything about GIS revolves around location with reference to a certain point on the earth. For this reason a discussion on projections are provided.

Projection is the process of projecting the round three dimensional surface of the earth with minimum distortion on a flat two-dimensional map. Due to the fact that locations on the earth are defined by a geographic co-ordinate system a projected co-ordinate system is based on a geographical co-ordinate system which is in turn based on a sphere or spheroid. One major difference between a geographic and projected co-ordinate system is that a projected co-ordinate system has constant lengths, angles and areas across two dimensions (Kennedy & Kopp, 2000).

In a projected co-ordinate system latitude and longitude locations are transformed to x, y co-ordinates that serve as values for a certain location on a grid, with the origin of the location at the centre of the grid.

However, projecting the earth is not as simple as projecting one's shadow against a wall. For instance the earth represents a round object containing a graticule drawn on it. By placing a medium of light inside the centre of such an object would project the graticule against the paper. In both cases, there is a form of transforming a three dimensional surface to a flat surface, but when it comes to projecting the earth, there are much more mathematical calculations included. These mathematical transformations are referred to as map projections.

One problem that usually accommodates projection is distortion. Distortion occurs along the edges far away from the centre of the earth. The further away from the centre of the earth the more the graticule gets distorted. Because the earth is now represented in two dimensions, distortion in the shape, area, distance or the direction of the data could occur. Many different map projections exist and therefore each projection has a distortion of its own and the distortion in some of the four characteristics (Kennedy & Kopp, 2000).

With any kind of projection the importance of the fact that no projection could ever represent the earth without some distortion could not be overstated.

There is not such a thing as a perfect projection, however a more suitable projection for an area does exist and the properties of each possible projection for an area should be considered before a projection could be used. Each map projection could show one or more but never all of the following characteristics: true directions, true areas, and true shapes. According to Silberbauer (1997), the three most commonly universal map projections are the Gauss Conform (Transverse Mercator) projection, Lambert Conic Conformal projection and the Albers Equal Area Conic. For the purpose of this study there were looked at four basic projections with relevant characteristics to the study at hand. These are: the Gauss Conform Projection, Albers Equal Area Conic, Universal Transverse Mercator and the Lambert Conic Conformal projection. The Universal Transverse Mercator is among the four because of its great versatility and accuracy on many projections.

Referring to Table 2.12 it is quite clear that any projection listed in the table could be applied for the current study and would be suitable for accurate data projection. However, due to the accuracy, versatility and minimal distortion that the Universal Transverse Mercator provides, it has been the projection that suits the requirements for the study more clearly (Lindenberg, 2011; Silberbauer, 1997; Kennedy & Kopp, 2000). Furthermore, the current QuickBird image (2008) that is used as backdrop reference for the CAD contains a spatial reference of WGS_1984_UTM_Zone 35S.

Following Table 2.12, section 2.8 provides a discussion on relevant case studies for the study at hand. The discussions of the case studies will provide information relevant to the current study and would indicate how GIS was applied to address the same challenges such as currently investigated.

Comparison of projections			
Projection	Typical scale	Application	Additional information
Gauss Conform (Transverse Mercator / Gauss conformal projection)	1:10 000	Orthophotos	Cylindrical projection. Used for local co-ordinate systems. Very similar to the UTM only with 1-degree wide column width. Distances are true along the central meridian or along two lines parallel to it and all distances, directions, shapes and areas are reasonably accurate within 15° of the central meridian.
	1:50 000	Topographical maps	
	1: 250 000	Topographical maps	
	1: 250 000	Topocadastral maps	
Lambert Conform Conic	1:500 000	Topographical maps	Planar projection. Used at regional scale. Especially appropriate for projects covering a large chunk of South Africa and need to measure areas or distances. Whenever there are worked with distance or direction Azimuthal Equidistant could be applied due to its measure of preserving shapes by sacrificing the accuracy of area.
	1:500 000	Administrative maps	
	1:500 000	Aeronautical maps	
	1:1 000 000	World aeronautical maps	
Albers Equal Area Conic	1:1 000 000	Maps of South Africa	Conic projection. Used at regional scale. Correlates well with the Lambert Conform Conic projection and its properties. Distortion of shape and scale are minimized in the region between the standard parallels. Used for small regions and countries – not for continents. Best results are for regions predominantly east-west in orientation and located in middle latitudes.
	1:2 500 000	Maps of southern Africa	
Universal Transverse Mercator	1:100 000	Topographic quadrangles	Cylindrical projection. Used for satellite imagery. Applied in areas where one is working with 6-degree wide columns around the earth (3-degrees-to-each-side-of-the-central-meridian). In other words, the earth is divided into 6 degree zones with a central meridian in the centre of the zone. Overall distortion of scale is minimized inside entire zone.

Table 2.12. Comparison between the different projections (Lindenberg, 2011; Silberbauer, 1997; Kennedy & Kopp, 2000)

2.8. Case studies

Case study 1

Water/wastewater authority geodatabase design and implementation

(Babbitt & Godfrey, 2003)

Background, problem and objectives

Water and wastewater infrastructure have not been available in digital format which prohibited the application of GIS in the water and wastewater management system. However, the Lehigh County Authority (LCA) located in Allentown was familiar with the advantages and capabilities of GIS. The objectives were to invest in a digital base map of the area for piloting efforts. Planimetric vector mapping and digital orthography were obtained. An approach towards the development of a GIS system had to be proposed. It was therefore important to start off with a clear indication of the broader picture for the development of a GIS system.

The GIS design phase

The GIS design phase consisted of different components that served as a structure for the proposed projects as presented to the Authority Board of Directors. The components of the GIS design phase are summarized. These were:

- a combined needs analysis and a GIS strategic implementation plan;
- recruiting a team representing each department in order to provide an inter-departmental buy in; and
- review the current pre-GIS environment:
 - review included documentation of the organizational structure and responsibilities, adopted policy documents, existing computer systems (hardware and software) as well as a complete inventory of the digital and hard copy data;

- hardcopy maps was limited to a schematic representation on small-scale index maps, a series of valve maps for each development and about 900 as-built development drawings. CAD drawings as well as a valve and hydrant database which were maintained in Microsoft Access and a home-grown Customer Information System were also included;
- key information and data gathering took place by means of interviews;
- priorities for applications have been implemented in order to prevent applying too many applications at once;
- following the prioritization of the applications a conceptual database design was undertaken in order to identify the thematic data layers to support the priority applications;
- a digital orthography and tax parcel data layer were developed by the local county government which was joined perfectly with the proposed plans for the project; and
- a detailed database design was necessary in order to develop a prototype application.

Even though the database design was limited to ArcView 3.2 the design was based on the ArcINFO arc/node topology. Line and node identifiers were used to differentiate between the various feature classes (distribution mains, hydrant laterals, valves, hydrants, fittings, manholes etc.). Detailed data associated with each feature type were stored in separate tables. A complete database picture of the features was created by joining a series of code tables which were created for the database features.

Summary and relevance to the study at hand

The drawings and plans were compiled in CAD and later translated to ArcINFO where coverages were built and additional attribution was added. A cost/benefits analysis was also executed. At a later phase the pilot coverages were imported into the geodatabase. ArcGIS were also used to georeference scanned images of the land base that were applied as a backdrop to the digitized infrastructure. The geodatabase also provided the benefit that the network topology was built right into the data model instead of being managed independently by an external application. The geodatabase design also proved to be less complex than originally expected.

Increased investment in GIS was performed due to the benefits of streamlining the work executed through it, which in turn resulted in improved customer service. The use of GIS also saved the company great amount financial expenses and emphasized the true ability of this software system.

Important concepts that have to be stressed when developing a GIS system:

- the implementation should be guided by a detailed and phased plan (steps);
- understand the technical and organizational impacts that the newly integrated GIS system might have on your organization;
- full implementation of the system should be done after a complete pilot and prototype conversion was executed;
- in case of enterprise wide application of the system it is very important to have all levels of staff involved in the design;
- get input of members of staff or target group into the system in order to assure appropriate values;
- good, open channel of communication is very important between staff members and consultants during all phases of the project; and
- stick to the plan but be flexible for smaller adjustments and the shifting in technological paradigms.

Case study 2

Challenges in developing a city GIS wastewater geodatabase system (Smith, 2010)

Background, problem and objectives

Municipal management in the City of Winfield Kansas provides good maintenance and provision of public utilities and safety services for its residents. The utilities that include the provision of electricity, natural gas, solid waste, potable water and wastewater services are supported and maintained with the application of GIS software. This was done by means of an SQL SDE multi-versioned database. GIS were applied as a medium for mapping and

maintaining the base map layers of the city and served as a medium for simultaneous editing of its utilities. With the attempt to integrate GIS in the daily operations concerning the service and maintenance of the city's utilities, the need to upgrade its aging wastewater and sewer treatment system were realized.

Due to the fact that the wastewater system of Winfield was totally neglected over the years, the need to upgrade the system arose. Even though the system were maintained through different formats, which included the use of an Auto CAD system with map booklets, a GIS shapefile system, paper log books and maps, these data formats were never updated. The wastewater systems were also created from existing drawings that were inaccurate due to an absence of accurate surveying and the lack of GPS technology. The result was a lack of adequate ground accuracy.

For these reasons the Winfield wastewater system was never used in the overall design and construction of the system. It was not maintained in a sufficient way nor was it able to generate reports or even capable of integrating data from different software applications and datasets. Furthermore, the data of the wastewater system did not exist in a relational or geodatabase system but rather in a "flat file" or "simple shapefile" system.

However, the core problem that gave way to the previous mentioned complications were the fact that no formalized attempt has ever been executed to identify the needs for the wastewater GIS system. The people that serve as the sources of information include the wastewater crew, the supervisors, directors and the operational staff. The problem had to be addressed and the solution started with the following objectives:

- to complete an analysis that looked at the requirements and the needs of the users;
- to look at the expectations of the wastewater department; and
- to investigate how GIS can be applied to contribute to the solution of the problem.

Data integration and applications in GIS

Integrating the laser fiche image document into the GIS system was recommended. This will include all the maintenance and tap information. The system should be opened to be used by the operational staff and the wastewater field crews for everyday use. In order to provide the possibility for future public use and city-wide access, the documented imaging would need to be linked up with the GIS. By maintaining the information within a database, the system was

created with the ability to track different types of problems and obstructions along the network lines in order to have these areas inspected.

Geodatabase design

The stages in developing a geodatabase are represented in the following summarized paragraph. Following the data collection stage and the review of existing data a geodatabase design were implemented. The different stages of the geodatabase design process are:

- the compilation of a “needs analysis” in order to generate an overall accurate indication for the eventual use of the GIS wastewater system;
- data collection included interviews, GPS collection, the examination of existing data which included paper documents, maps, Auto CAD files and databases, GIS shapefiles, Microsoft Access, spread sheets, laser fiche documents, AS400 applications, construction and as-built drawings;
- with regard to the accuracy, data of the underground utilities have been collected on a standardized co-ordinate system for the city; 9NAD 1983 State Plane Kansas South FIPS 1502, GCS North American 1983 Datum. The importance of a standardized co-ordinate system is emphasized; and
- according to existing and collected data, database attributes were build in GIS for each layer.

According to the specific requirements for the system a selection of applicable thematic feature layers together with their attributes that would be represented in the geodatabase were chosen. “Future collection” and data “integration” were some of the important factors that had to be stressed. In order to provide a structured way for future collection of maintenance data and data integration, different attributes were created accordingly.

The city was divided in maintenance sections to assist with the collection of data, the database design and classification of maintenance information. A database was created shortly after the GPS collection for the features were completed and existing data sources were reviewed. Data collection and assessment therefore formed the first step in the creation of the database. Data editing followed soon after the data were in place and existing wastewater lines were connected to the manholes by using the snapping operation. The use of unique identifier numbers were created and applied for different facilities which also contained certain characteristics of the facilities it represents. The following procedure was

to create database attributes in GIS for each layer that was identified in the user requirements analysis. Attribute tables of features such as manholes, forced mains, cleanouts and connected lines were updated from existing data. Remaining data fields were updated from existing data originated from construction and as-built drawings, log and maps books and various discussions held with field personnel.

It should be mentioned that the system were created with programming software, Visual Basic Application (VBA) code. It used GIS shapefiles, Access databases, and MXD project files. This system was designed to provide the engineering staff with the ability to update information about any operations executed on the system on any type of feature. Any adjustments to the system are automatically updated. The new system provided an interface for the user with the option to indicate which action would be performed on the system. ArcMap 9.3.1 served as a medium to enter field maintenance information by specifying different actions to be executed and by making use of the unique identifiers to locate the point of maintenance. A prototype of the system were first fully tested and executed until it was ready to be migrated into an ArcSDE multi-versioned database where further adjustments according to the new environment were executed.

The outline of the database creation in this instance has been briefly discussed with the emphasis on only a few important aspects. However, it is important to stress the fact that a logical order does consist behind each database design and should serve as a detailed guide for effective implementation of the database.

Summary and relevance to the study at hand

The case study provided a brief overview on the procedures and the solution found to the problem of a water system. A few comparisons could be drawn between the case study and the current study at hand. Each project is unique in its own way but it is important to have a frame of reference to serve as a guide for the creation of a database for any system. A very important aspect of this study is the principle of communication between the working staff that had to have a correct indication of the requirements for the utility system.

Important aspects of this case study that should be emphasized:

- the problem were identified;
- a needs-analysis were executed;
- a reference scale for the data were identified;

- the study area was divided in smaller manageable parts (zones);
- the layers together with their attributes were identified;
- the possibility of future applications were identified;
- from the “needs-analysis” functional system requirements for maintenance purposes were created;
- data were integrated within GIS;
- a problem tracking system were created in order to trace any problems along the network link;
- data collection and data storing were performed by means of GPS, CAD drawings, paper records, PDAs and laptops;
- a database design followed after the data were collected;
- unique identifier numbers were created for each facility;
- the GIS system did not function in isolation of other software systems; and
- a prototype was first executed.

Case study 3

A GIS data model for enhanced navigation in urban environments (Mandloi, 2007)

Background, problem and objective

This case study represents a project executed for the North Campus of the State University of New York at Buffalo. The objective was to create an object-oriented data model that would represent a multi-modal urban transportation network. The model would be designed for navigation purposes and various kinds of network analyses in urban environments. It would serve as a medium for viewing movement inside and outside buildings, as well as to simulate and model different scenarios and navigation procedures or emergency planning. In this case study the 2D and the 3D GIS environments have been applied. The 2D GIS environment consists of a multi-modal data model for networks outside of the building while the 3D environment consists of a network for the movement inside buildings on more than one level

and focused more on the visualization aspect of the study. The dualistic nature of the study, containing the 2D GIS for walkways, streets and public transit and the 3D GIS environment for high-rise buildings and transportation provides a good indication of the integration between the two environments.

Due to the fact that both environments consists of a network structure, the network data model represents a very strong component of techniques applied to store topological relationships of connectivity between network entities. These topological relationships of connectivity are supported by the node-arc data model that serves as the foundation for storing these relationships. However, some complications arise whenever the node-arc model is applied to the 3D nature or non-planarity nature of a network and hence prevents the complex reality in some situations. Due to the 3D nature of the buildings, the source of elevation for those 3D entities are not derived from the X, Y co-ordinates of the co-ordinate system used for the project but are used as attributes in the system. It is therefore possible to view the vector entities in 3D but not have any topological relationships defined for it.

The project was executed by applying commercial off-the-shelf GIS software, ArcGIS. With the use of this software a 3D path finding application called Network3D were created. The data model was developed with the integration of sources such as aerial photos, CAD drawings and paper maps.

The geodatabase design

Three stages of database design were applied in the case study. These were:

- the conceptual model;
- the logical model; and
- the physical model.

Furthermore the use of feature datasets, feature classes, subtypes and network datasets were applied in the geodatabase. The approach towards representing different floors in 3D GIS was executed by storing all the data of all the buildings and all the floors in one single feature class. The different floors were distinguished by placing each floor in a separate subtype. In order to connect all the sections and features of the area outside the buildings a multi-modal network dataset have been created.

ArcGIS Network Analyst was applied in order to perform network-based analyses such as finding the shortest route and optimal paths. The data model was used to store 3D connectivity. 3D visualization and interaction of 3D routes within the buildings were made possible with the use of ArcGIS 3D Analyst.

The following results derived from the study were:

- finding least effort route between location in a building;
- a least effort route between two locations and between the buildings;
- determining a least effort emergency route between two locations in a building; and
- an all-inside least effort route between two buildings connected by an enclosed walkway.

With the use of an object-orientated GIS data mode, navigation in urban environments was applied. This was done in accordance with multiple transportation modes. Navigation inside buildings was made possible with the use of the object-oriented GIS data model. The 2D network modelling capabilities of the current ArcGIS system provided the ability to model 3D network connectivity.

Summary and relevance to the study at hand

The case study provided a good overview of the integration between the 2D world and the 3D world as well as their applications. After various applications it was decided to model the floor levels in one feature class instead of creating a feature class for each one.

A strong component of the network dataset and network structure was resembled and by applying ArcGIS 3D Analyst, the 3D visualization and interaction of 3D routes within the buildings were made possible. It was proven that topology is not yet suitable for the specific 3D application and that alternative ways, such as using a network data model were better suited to model the 3D network. The geodatabase design steps as provided by Arctur & Zeiler (2004) were successfully applied and proved to be a good starting point for the design of the geodatabase in support to current study with the focus on the 3D environment.