

Chapter 2: Literature review

This Chapter describes the physical and logical application of computer networks in the real world. Chapter 2 then compares management tools which can be applied to manage the computer infrastructures. The decision was made to implement a GIS system, and therefore a large part of the Chapter explains the basic concepts of GIS, with the main focus on a geodatabase. The geodatabase concepts are described in detail and three geodatabase types are compared in order to make an informed decision as to which geodatabase type will be best suited for this study. The chapter then compares three geodatabase design methods, in order to utilize it as a guideline when designing this data model. Finally the chapter examines case studies, which are similar in some areas to this study.

2.1 IT-infrastructure application

2.1.1 Defining an IT-network

Networks connect computers to one another, through a combination of hardware and software, sharing information and peripherals in an economical and efficient manner. The computers or other devices which are connected via the network are known as nodes. The network allows geographically distant computers to share information with one another, as well as allow multiple users to share one or more devices. A network can also be defined as a collection of independent computers, interconnected by a single technology. Two computers are described to be connected when they are able to exchange information (Tanenbaum, 2003; Naugle, 1994).

Different types of networks exist. The smallest is the local area network (LAN), which is normally privately owned networks within a single building or campus up to a few kilometers in size. They are widely used to connect personal computers and workstations of a company which share information and resources. A metropolitan area network (MAN) covers a city, such as a cable television network. Another type of network is a wide area network (WAN), which spans a large geographical area, usually a country or continent. These are normally

owned by a service provider such as a telephone company. Networks can be connected to one another through machines called gateways, which translates information in terms of hardware and software, thus creating the internet. The internet essentially connects different networks to one another, creating a global connected environment (Tanenbaum, 2003).

Due to the fact that the Potchefstroom campus implements a LAN, only the LAN network will be examined, in order to create a greater understanding of how the network is employed, what hardware is used, and how it is maintained.

2.1.2 Local area network topologies

In terms of network layout and connectivity, the term topology describes the physical configuration of a network. There are 4 main topology types of IT networks. It must be noted that these types are purely theoretical and used as a foundation model to describe the networks. In reality, networks do not often fall in single types. It is the nature of networks to combine different technologies and architectures to form hybrid designs (Naugle, 1994; Serumaga-Zake, 2006).

- The Star topology is also known as a wiring hub. As seen in Figure 2.1, it consists of a central hub (switch), which is connected to all the nodes. In this context the nodes represent the device, which can send and receive information, and is part of the network. The topology has no single failure point which will affect the network, except for the hub itself. This is one of the most popular topologies, and allows for efficient network management (Tanenbaum, 2003; Naugle, 1994).
- As seen in Figure 2.1, the Ring topology encloses all nodes in a loop and considers them as repeaters. A repeater can both receive and transmit information. If a station receives information which is not intended for it, it re-transmits the information downstream. If a single station is down or removed from the network, it can affect the whole network (Tanenbaum, 2003; Naugle, 1994).

- The Bus topology consists of a single shared length of cable, to which all nodes are connected (Figure 2.1). This topology is also known as a linear topology. The network terminates at the endpoints and has no closed loop. If the cable is damaged at any point, all the attached nodes are removed from the network. The bus topology is normally used for Ethernet networks (Tanenbaum, 2003; Naugle, 1994).
- The Tree topology is a generalization of the bus topology. As seen in Figure 2.1, this topology has a hierarchical undertone. It has a root point (normally a central hub/switch), from which all branches stem. There are no loops in the network, which means there is only a single data path to any endpoint on the network. If a network error occurs at a certain node in the network, all the dependant nodes down the hierarchy are disconnected from the network (Tanenbaum, 2003; Naugle, 1994).

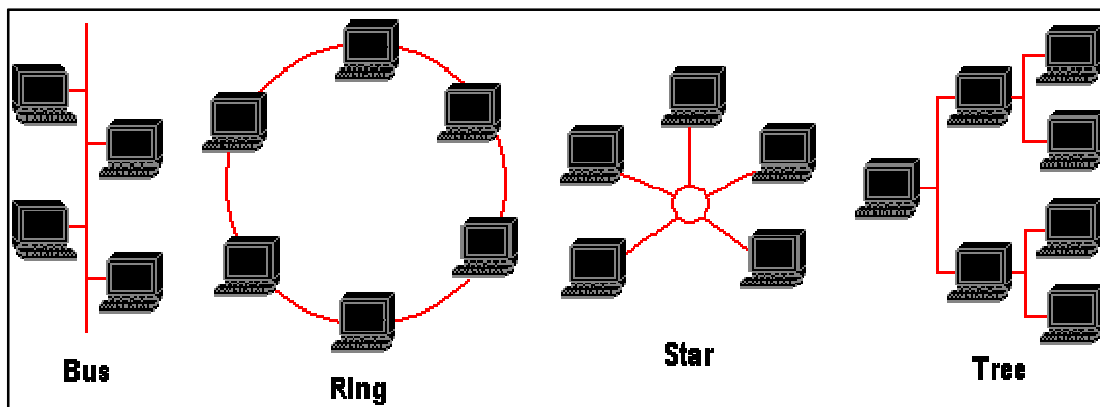


Figure 2.1: Network topologies (Anon, 2011)

When designing a network, it is very important to include precautionary measures to prevent a total network failure in the event that certain elements cease to work. An example of this is to ensure that a back-up central hub is in place, should the primary hub go down, when a star topology is implemented. Another example, when designing a ring topology, would be to ensure that when one station is terminated, the data flow can continue in the opposite direction (Tanenbaum, 2003; Naugle, 1994).

2.1.3 Cable topologies for buildings

There are various ways of wiring a building. As mentioned above, some network topologies do exist for the general network, but other topologies, specifically created to help with the wiring design inside of a building also exist. The most simple building cable topology is the linear building topology. This schema consists of a single cable, snaked through each room, where each device connects to it at the nearest point. This is a bus topology (Naugle, 1994).

The backbone topology consists of a thick cable running vertically through a building (backbone). Thinner cables connect to the backbone on each floor via repeaters and spreads throughout the floor to the nodes. A repeater is a device which receives the signal, amplifies it, and then retransmits it to the destination. A repeater is also known as a concentrator. Repeaters overcome the cable length limit and the maximum attachments limit by extending the physical topology of the network (Naugle, 1994; Tanenbaum, 2003; Serumaga-Zake, 2006).

The most popular building topology is the tree-network. This is a hierarchical network where hubs are used to redirect the transmission to the right destination down the hierarchy. The transmission may pass through a couple of switches up or down the hierarchy, but ultimately each destination has one path (Naugle, 1994).

2.1.4 Network hardware

A network consists of a variety of different connected hardware types. One of the most important of all the hardware types is cables. Cables are responsible for connecting different types of hardware and devices over long distances. The term "Ethernet" refers to the cables in a network. There are mainly five different types of cables used in networks. The implementation of each largely depends on the geographical scale of the network, the potential data carrying capacity needed and the number of connected features. It is common for a network to implement the use of more than one type of cable. The five most commonly

used cables are depicted in Table 2.1 (Naugle, 1994; Tanenbaum, 2003; Serumaga-Zake, 2006).

Cable type	Standard name	IEEE name	Maximum length (m)	Maximum speed (Mbps)
Multi-mode optical fiber	1000 BASE SX	IEEE 802.3z	550	1000
Single-mode optical fiber	1000 BASE LX	IEEE 802.3z	10000	1000
Category 5e Unified twisted pair (UTP)	100 BASE T	IEEE 802.3u	100	100
Thick coaxial cable	10 BASE 5	IEEE 802.3	500	10
Thin coaxial cable	10 BASE 2	IEEE 802.3a	185	10

Table 2.1: The 5 most commonly used network cables (Naugle, 1994; Tanenbaum, 2003; Serumaga-Zake, 2006)

The thick coaxial cable (10BASE5) also known as thick Ethernet was the original network cable, but since the latest cable releases, have become obsolete. It had a maximum data carrying capacity of 10 Mbps. More recent cables are able to cover greater distances and carry much more data. The next to follow was the thin coaxial cable. It was much cheaper and easier to install than its predecessor, but the maximum length of a segment was only 185 m and could only accommodate 30 machines. Detecting cable breaks in one of the coaxial cable systems was hard, which drove the cable industry to develop a new kind of wiring pattern (Naugle, 1994).

This new system consisted of every machine connecting a cable to a central hub, in which they are all electrically connected, as if they were fused together. Due to the fact that each node has its own cable, each one can easily be added or removed, which simplifies break detection vastly. The cable type for this design is unshielded twisted pair (UTP). This is the same type of cables used by telephone companies. UTP cables can be sub-divided into categories. Category 3, 4, and 5 cables can transfer data at 10, 16 and 100 Mbps, respectively (Naugle, 1994; Tanenbaum, 2003).

A fibre optical cable is the final cable type. The cable consists of ultra thin fibre of glass inside the cable, angled at such a degree that when a pulse of light is injected into the cable, none of the light escapes. The transmission medium has both a light detector and light source on either side of the cable. The light source accepts electrical binary pulses as input, converts it into light pulses and sends it along the fibres. The detector generates an electrical pulse when light falls on it, and converts it to an electrical pulse as output. The expensive fiber optic cable segments can stretch for kilometers and is therefore mostly used for long distances between buildings. There are different types of optical fibre cables, including the multi-mode optical fibre cable and the single-mode optical fibre cable. The multi-mode optical fibre has a maximum length of 550 meters, while the single-mode optical fibre has a maximum length of 10 kilometers (Naugle, 1994; Tanenbaum, 2003).

A repeater is another hardware type that is occasionally present in a network. As mentioned before, repeaters extend the physical topology of the network by extending the number of potential attachments and maximum length of cable segments. For instance, four repeaters on a 10Base5 cable will lengthen the maximum length from 500m to 2500m. In short, repeaters make it possible for a greater number of nodes to connect to a single cable. A cable segment may have a maximum of four repeaters before the signal strength weakens (Serumaga-Zake, 2006).

There are different types of switching devices; each with its own diverting and collision management parameters, but in essence performs the same task. Switching devices can interrupt a current in a data conducting cable, and redirect the transmission to another conductor, thereby expanding the network into different branches. These devices include repeaters, hubs, switches, routers and gateways. For the purpose of this study all switching devices will be referred to as switches. Some other devices which take part in the network are:

- Ports (also known as Terminals): These are all the sources and destinations in the network.
- Hosts: Hosts are large computers serving many users, providing computing capabilities or access to a database. They are also known as servers or sources.
- Multiplexors and concentrators join the traffic of low speed lines into a single stream, which can use a higher speed line (Serumaga-Zake, 2006).

It is not mandatory for all of the above mentioned hardware to be present in a network. Most networks, such as the one used as the main focus of this study, implement a combination of the various hardware types (Serumaga-Zake, 2006).

2.1.5 Potchefstroom campus IT network

The campus IT-infrastructure fundamentally consists of three types of network elements, switches, cables and network ports. The campus has four distribution star topologies spread around on campus, which are connected to one another. Each star consists of a number of switches. One of the stars serves as the main server of the campus. The main server will be referred to as the source for the remainder of this study. If the distribution stars need to communicate with one another the data bundle will have to travel via the source. The distribution stars are connected to the source through a single-mode fibre optic cable. Each star serves as a gateway to the source for a number of switches inside of buildings of a specified area. For example, in this study the source connects to the library distribution star, which in return connects to the switches in buildings E4 and E6 (Buys, 2010).

Each building has at least one IT utility room, which contains the switches for the building. The room contain a number of switches depending on the number of network ports. Each switch can only serve 24 network ports. One of the switches is connected to the distribution star through a single-mode optical fibre cable. Each switch is connected to the adjacent switch through a multi-mode optical fibre cable (Buys, 2010).

The switch is connected to each individual network port through a Category 5e UTP cable (Cat 5e). For instance, if a switch serves 24 network ports, there will be 24 Cat 5e cables connected to the switch. The network which exists between the distribution star and the network port is a tree topology. The campus infrastructure implements a combination of the star- and tree topologies (Buys, 2010).

Such a large scale network is constantly under immense pressure. Infrastructure constantly breaks down. There is presently a lack of available information about the geographical

location of infrastructure elements. Where there is information available, it is primitive (such as hardcopy paper maps) and not always up to date. Implementing a management tool could benefit the utility management and maintenance technicians. A system is needed to digitally store, manage, update, analyze the information as well as visually display the geographical location of the elements (Buys, 2010).

2.2 GIS vs. alternative information management methods

This section lists the benefits of implementing a GIS, as well as describes examples of how GIS can be implemented. It also compares GIS software to alternative information management methods which can be implemented as a management tool for the IT network infrastructure of the campus. The aim is to determine the benefits and shortcomings of each in order to make a good decision as to which information management tool will be implemented for this data model.

2.2.1 Application and benefits of GIS

ESRI (2008a) defines GIS as a system which combines database management tools with mapping software to collect, share, edit and organize many types of information. Athanasios et al. (2009) defines a GIS as an application which combines spatial interpolation and a spatial database (a database containing longitude and latitude parameters) into a single system that allows an operator to interact with both aspects simultaneously. This allows the data to be linked to a graphical representation. The ArcGIS pipeline data model (APDM, 2004) defines a GIS as a powerful tool to capture the location and extent of features (or assets); schedule their maintenance or replacement; identify and mitigate potential risks; identify the impact of failure; design expansion activities; manage customer queries and analyze environmental issues.

The main goal of GIS software is to help the user solve problems and answer questions regarding his/her data, by analyzing and representing the data in an easily understood fashion. GIS offers ways to map locations, quantities, densities, specific areas and

environment change. For the last thirty years a diverse selection of industries have used GIS to store, manage and analyze their data in order to optimize production, management and distribution. The business sector, government, educational and scientific sectors, environmental management and conservation, natural resources and utilities are all industries which incorporate and apply GIS software. GIS can essentially be used for a very large number of applications. For example, in the utilities sector, GIS is used for power management, gas resources and distribution, telecommunications and water and wastewater (ESRI, 2010a; ESRI, 2005).

GIS has improved organizational integration between different departments. By creating a shared database, one department can benefit from the work done by another. This enhances inter departmental sharing and communication. Data can be collected once and used many times. GIS is in many instances used as a tool to achieve better decision-making, although GIS is not an automated decision-making tool, it is often used in order to represent data in such a way that illustrates potential outcomes (ESRI, 2010a).

Infrastructure management is a complex field. Management incorporates the installation, monitoring, upgrading and repair of the specific utilities. In order to efficiently accomplish these goals, having the correct information about the infrastructure's spatial location as well as its specific attributes at hand is vital. GIS offers a method which stores the spatial location and attributes of a feature in a central place. The GIS also contains information about validation rules, feature domains and relationships with other features, as well as offer analysis tools. GIS can not only be used for a wide variety of industries (such as the military-, health care- and economic sectors), but can also be used for a range of utilities (Lewis & Ogra, 2010). Table 2.2 lists a combination of benefits described by ESRI (2010a) and the APDM (2004).

Although GIS offers a number of various benefits, it is not the only geographical information management tool. The following section firstly compares GIS to the original spatial representation method; paper maps. The section then compares the differences between GIS and Computer-Aided Design (CAD) systems and describes the importance of integrating the two. The importance of integrating GIS and Database Management Systems with one another is also highlighted.

Benefits of GIS	
1	GIS is a computerized map which integrates features (points, lines and polygons) with attribute information to build a map-based information system.
2	A GIS can perform geographical queries and analysis
3	GIS offers the ability to add additional criteria to a query
4	GIS manages assets in terms of recording information about the assets such as its value, condition and location.
5	GIS can re-combine existing data to make new data. Data in different layers can be merged together intersected with one another or overlain on top of each other. This allows for powerful visualization patterns in a map, which could not be done in a normal database.
6	GIS improves decision-making by providing analyses of spatial and non-spatial elements.
7	GIS is part of the enterprise. When the databases of different systems, such as asset management systems, or customer information systems are linked through foreign keys (identifiers) an Enterprise is created. GIS through its unique database forms part of the enterprise.
8	GIS provides advanced data structures for optimal analyses of the data. These data structures include network analysis (route planning), spatial analysis (raster analysis), 3-D analysis (terrain visualization), geometric calculation between features (intersections, overlays and adjacency) and topology (relationships between features in space).
9	GIS maintains complex data accurately. When huge quantities of data are being handled, faulty or missing data can easily be identified and edited.
10	GIS software provides a variety of interfaces such as desktop, server tools, internet web browser, depending on the needs of the business.
11	GIS generates real maps. Maps show co-ordinate positions and vector directions as well as features in proximity to each other. Maps improve communication immensely; it creates a visual image which is much easier to understand than numeric descriptions. Appleton & Lovett (2003) states that: "Visual communication is an increasingly common part of environmental decision-making, being used as a 'common currency' to facilitate dialogue between policymakers and non-experts, increase understanding, and thereby improve the decisions made."
12	GIS provides economic benefits. GIS benefits a business or project by streamlining workflow and business processes. For example, when an assets location in the field, attributes and proximity to other features are known, it saves time, helps to schedule regular maintenance and helps expansion planning, respectively.

Table 2.2: Benefits of GIS (ESRI, 2010a; APDM, 2004)

2.2.2 Differences between CAD and GIS

CAD has many definitions, depending on the industry that it is applied to. Rath (2007) gives a generic definition of CAD in its simplest form as a process where design data is converted into digital format, which enables the data to be handled easier and quicker. Industries which make use of CAD include Engineering, Architecture, Fashion, Textile and Graphics. The basic entities used in CAD are points, lines and polygons, but these entities have been adapted to create circles, ellipses, parabola, hyperbola, mesh as well as some 3D objects such as a sphere, rhombus, cube and cylinder. CAD drawings are created according to dimensions given in a hard copy or dimensions mentioned in an image or raster file. According to the study done by Morgan (2004), CAD systems are often used by engineers and architects to design building site plans. CAD systems offer a variety of tools which aids the designer to achieve his or her objectives.

Peachavanish *et al.* (2006) describe the differences between CAD and GIS in terms of two categories: reference system and space. CAD software has a local coordinate system which is only applicable for the current site whereas GIS has a geographical coordinate system, which references objects with its position relevant to the earth. CAD files have both 2- and 3-dimensional environments. According to Martin (2008), AutoCAD created AutoCAD Map 3D as an independent 2D GIS application, while AutoCAD Civil 3D was used as a 3D viewer. ArcGIS can perform spatial analysis in a 2-dimensional environment (ArcMap), while viewing 3D data in the 3D Analyst extension (ArcScene and ArcGlobe). The latest ArcGIS 10 software has however started to develop some degree of analysis in 3D, which makes it possible to do connectivity examinations among network features in a 3D environment. According to Murphy (2004), as-built CAD information historically always lacked a coordinate system, was never to scale and was always unstructured. GIS however offers editing tools, which allows the user to edit and scale the CAD data so that it corresponds with GIS data (ESRI, 2011).

ESRI (2001) lists some of the restraints for CAD software. CAD software lacks a database environment and has to be integrated with an external database management system (DBMS). CAD software also lacks the ability to do spatial analysis on the data. Spatial analysis is when a user combines different independent data in order to create entirely new

information. GIS offers integrity rules through topologies, which specify the spatial connectivity rules between geographically integrated features. CAD software also has no topological information. CAD software does not have layers in the same sense as GIS does. GIS layers each represent a different feature and can be turned on and off by the user, whereas CAD software shows all the different features on one display. When CAD data is viewed in GIS (ArcCatalog), the CAD file is divided into different layers. One layer is created for every feature shape, such as one for all the polylines; one layer for all the polygons; and one layer for all the points. There is no distinction made between the different features except for the design legend, which can lead to confusion. For example, a CAD polyline layer may depict the roads, railway lines and electrical lines, all in the same feature class. In GIS however a feature class is created for each separate feature, which can be easily activate or deactivated independently (ESRI, 2011).

According to Pu & Zlatanova (2006), GIS and CAD models are getting closer than ever before. CAD was originally created as a designing tool, with the primary aim on effective 3D visualization and accurate editing tools, within a local coordinate system. GIS was designed to represent objects according to the real world. Accordingly GIS used geographic coordinates, maintained shapes such as lines, points and polygons, as well as containing the shape's corresponding attributes. GIS aimed to replace all the methods that were used to do analysis on paper maps (Pu & Zlatanova, 2006). Presently the two types of software are becoming more similar. Engineers using CAD, are starting to do GIS-like analysis as well as using complex hierarchies of attributes in their designs, while GIS users demands realistic 3D visualization such as in CAD.

2.2.3 The importance of integrating CAD and GIS

The importance of integration between CAD and GIS is highlighted by the fact that their designs are evolving in the same direction. Drawing designs created by engineers and architects in CAD often have to be integrated into a municipal GIS. Murphy (2004) describes one example as the town of Truckee, California. The municipality's electrical system design work was completed in-house using GIS, while electrical and water schemas were created using CAD by independent architectural and engineering companies. The solution was to build an enterprise GIS and integrating the two formats. Peachavanish *et al.* (2006) states

that infrastructure managers often rely on both CAD and GIS for making decisions when engineering designs are being implemented. GIS is used to deal with problems with geospatial data, while CAD is used to deal with problems concerning design errors. CAD and GIS integration generally implies using CAD to record detailed design information about an object or structure, while using GIS to capture information about the surrounding area and doing location-related analysis at different scales. In the proceedings of the Southern California CAD summit of 2008, King (2008) listed areas where CAD designers can make changes in their methodology in order to simplify data integration. These changes include the conversion of drawings into different coordinate systems; transforming images for alignment; and using the “Drawing cleanup” tool to eliminate duplicates and other typical errors that result from scanning and file conversion.

Although the integration between CAD and GIS holds many advantages, some problems appear which makes integration difficult. According to Murphy (2004), these problems include the project-based coordinate system used by CAD compared to the geographic coordinates used in GIS. In order to align types of data, a singular geographic coordinate system needs to be applied to all data types. Other CAD data problems mentioned by Murphy (2004) are the cryptically named layer structure and that the data is often not drawn to scale. According to Peachavanish *et al.* (2006), the problems due to integration can be entitled to heterogeneities in information. These heterogeneities can be divided into three groups. First is the syntactic heterogeneity, which are the differences in syntax between the different software types. Structural heterogeneity deals with the differences in languages, interfaces and representation schemas. Finally semantic heterogeneity refers to the differences in the intended meaning of entities and concepts, for example the word “touch” may have a different meaning in a topology than in other software. In order to fully integrate these software types one has to bridge over all the differences between them.

2.2.4 GIS integration with database management systems (DBMS)

In terms of facility management, utilities started using GIS as early as the 1980's. These GIS applications were called AM/FM (automated mapping/facility management) and only ran on high performance computers, storing geographical features in flat files and attributes in relational databases (Harder, 1999).

Mapping-graphics which is linked to a large database is another advantage of using a GIS. Cowen (1988) described a database management system (DBMS) linked to mapping graphics, at its simplest form, as an electronic filing cabinet that supports selection procedures, sorting and queries, while the maps and graphs are just specialized output functions. Dempsey (2000) defines a simple DBMS as a system which provides input, retrieval and storage for data. GIS is able to combine with a variety of database management system software. An enterprise GIS is multi-user GIS, which integrates with an independent DBMS. In his study, Pacurari (2002) lists the advantages of integrating a geodatabase with a DBMS, which include: the easier integration of spatial data with other organizational data; an extended database size; and support for the larger number of users required for a successful enterprise GIS.

One example of the integration between GIS and DBMS is ArcFM (facility manager) in which all the utility's data is stored in a single relational database management system. This is made possible by the software's spatial database engine (SDE), which enables data to be stored, managed, and retrieved swiftly from leading independent DBMS (Harder, 1999).

When GIS is compared to alternative information management systems it becomes evident that it offers an accurate spatial representation of the real world. A GIS possesses genuine projected coordinate systems, spatial symbolism of real objects, and easily integrates with other software types; it also stores the data in a geodatabase which offers a variety of analysis tools. A GIS system can also be easily edited, updated or exported. For these reasons, GIS has been chosen as the software type to be used to develop the pilot data model for this study.

2.3 An introduction of GIS concepts

The following section describes the fundamental concepts of GIS. It illustrates the principles of coordinate systems and how it is applied to GIS. The basic elements of the GIS 3D environment are also defined, after which the section describes how spatial and non-spatial data is represented, as well as the data models for storing the data. Finally the section

compares three types of data storage models to one another and chooses the one with the most benefits.

2.3.1 Coordinate systems

One of the most important concepts in GIS, and cartography as a whole, is map projections. Although the earth is a spherical shaped planet, it is however not perfectly round. The term “Geoid” describes the true form of the earth. This section summarizes coordinate systems, in order to create an understanding to why a certain coordinate system was chosen for this study (Zhang, 2005; Delmelle, 2001).

The closest mathematical shape to the earth is a spheroid (also known as an ellipsoid). The most common method for assigning a surface position is by using a grid formation. The North-South lines are known as meridians and assign the longitude location of a point. The East-West lines are commonly referred to as parallels. The parallels give the latitude position of a location. When both the latitude and longitude values are known, it is possible to detect the location on the grid. There are a total of 360 meridians and 180 parallels (Zhang, 2005; Kennedy & Kopp, 2000; Delmelle, 2001).

Due to the fact that the geoid is not perfectly spherical, there exist some inconsistencies between the geoid and the spheroid (see Figure 2.2). The differences in height between the two are known as the ellipsoid height. At places on the surface, the spheroid and geoid do not have any height differences (where the geoid and spheroid intersect) these locations are known as a geodetic datum, or datum for short. A datum serves as a suitable origin for stated coordinates. A datum consists of three-dimensional Cartesian axes, which allows positions to be described in terms of latitude, longitude and ellipsoid height (Zhang, 2005; Kennedy & Kopp, 2000).

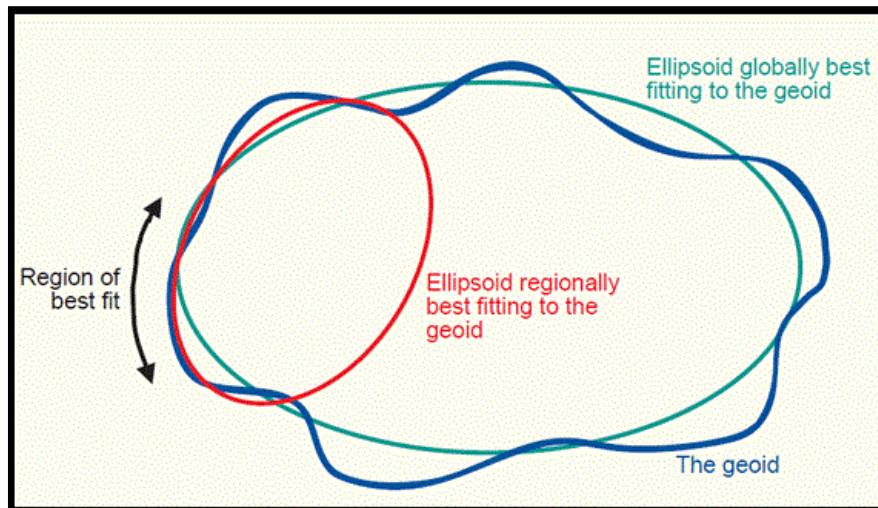


Figure 2.2: Differences between an ellipsoid and the geoid (Knippers, 2009).

According to Model Maker Systems (MMS, 2011), the first datum implemented in South Africa was named the Cape Datum and was based on the Modified Clarke 1880 ellipsoid. On the 1st of January 1999, South Africa's official coordinate system shifted from the Cape Datum to the Hartebeeshoek94 Datum, which is based on the World Geodetic System 1984 (commonly known as WGS84). The central point of this system is the co-ordinates of the Hartebeeshoek Radio Astronomy Telescope. If the same point was represented on both the WGS84 datum and the Cape Datum, the locations of the point for the respective datums would be approximately a distance of 300 meters apart. In South Africa the locally based datum Hartebeeshoek94 was founded upon the WGS84 datum and the differences may virtually be omitted (MMS, 2011; Wonnacott, 1999).

A coordinate system constitutes a grid which covers the surface of the earth, as well as a datum, in order to accurately locate a position on the surface of the earth. There are two types of coordinate systems; a geographical coordinate system and a projected coordinate system. A geographical coordinate system makes use of latitudes and longitudes to define a position on the surface of the earth. This type of system is used when the location of an object, relevant to the surface of the earth, is the main focus (Kennedy & Kopp, 2000).

A projected coordinate system represents the earth's shape as on a two-dimensional platform. Even though a projection portrays a section of the earth on a flat surface, projections are distorted from the original shape. It is impossible for any 2D projection to accurately represent the surface of the Earth perfectly. Representing the earth's surface in

two dimensions distorts the shape, area, distance, or direction of the data. There are different types of projections, the three most common types are the cylindrical projection, conic projection and planar projection and can be seen in Figure 2.3 (Kennedy & Kopp, 2000; Delmelle, 2001).

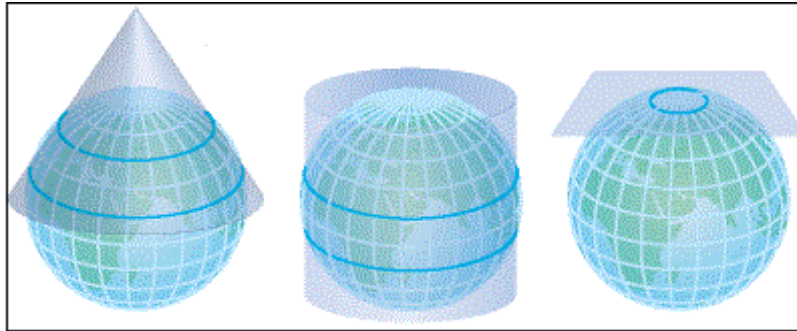


Figure 2.3: A conic projection (left), a cylindrical projection (middle) and a planar projection (right) (ESRI, 2011).

A projected coordinate system is used for mathematical calculations on the surface of the earth, such as the size of an area (square measurements). There are different types of projected coordinate systems, each suitable for a certain type of project. Projection types' functionality is determined by its representation of a feature's shape, area, distance and direction (Kennedy & Kopp, 2000).

In this study a projected coordinate system was implemented for the feature dataset in order to perform spatial analysis on the data model. Three projected coordinate systems were compared to one another in terms of their shape, area, direction and distance properties, as well as their limitations, as summarized in Table 2.3.

The Albers Equal Area projection is a widely used coordinate system (Figure 2.4) and is a conic projection. The Albers Equal Area projection is most effective on an area that is largely east-west in orientation. Total latitude range from north to south should not exceed 30-35 degrees. This projection type is commonly used on local areas instead of areas of continental extent (Kennedy & Kopp, 2000).

Projections	Projection type	Shape	Area	Direction	Distance
Albers Equal Area	Conic	Not fully conformal	Accurate	Accurate along the standard parallels	Some distortion
Transverse Mercator	Cylindrical	Conformal	Some distortion	Accurate	Some distortion
Universal Transverse Mercator	Cylindrical	Conformal	Accurate	Accurate	Minimal distortion

Table 2.3: Projection comparison summary (Kennedy & Kopp, 2000)

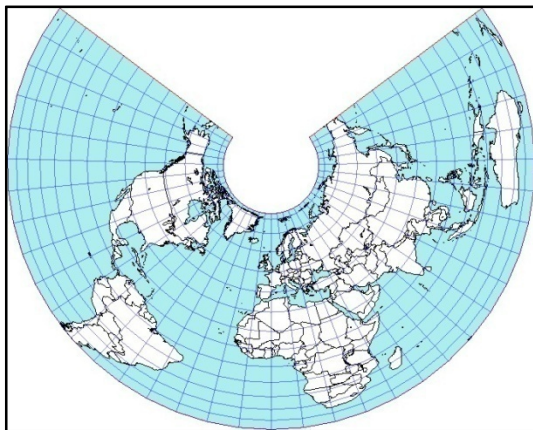


Figure 2.4: The Albers Equal-Area projection (Aquarius.Net, 2011)

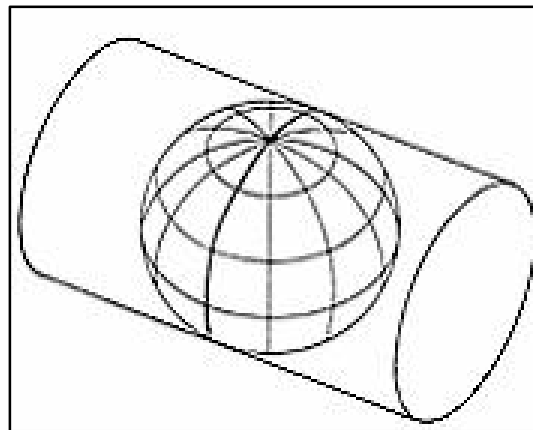


Figure 2.5: Cylinder for a Transverse Mercator projection (Fekete Associates Inc., 2011)

The Transverse Mercator projection is also known as the Gauss-Krüger projection, and is based on the Mercator projection. This is a cylindrical projection where the cylinder is along the meridians instead of the parallels (Figure 2.5). The projection is best suited for areas which stretch north-south (Kennedy & Kopp, 2000).

The final projection type is the Universal Transverse Mercator, also known as UTM. This projection is a specialized application of the Transverse Mercator: the world is divided into 180 zones spanning 6 degrees of longitude each; 60 north and 60 south zones. Zones N1 and S1 starts at the -180° W meridian. The division between the north and south zones occur at the equator. Although the UTM is only valid for the specified zone, it is very

accurate in terms of displaying shape, area, direction and distance properties. Similar to the Transverse Mercator the UTM also is a cylindrical projection. UTM northern zones are limited to the 84°N parallel, while the southern zones are restricted to the 80° S parallel. By dividing the world into zones, the UTM creates a very accurate local projection for each part of the earth's surface (Kennedy & Kopp, 2000). Figure 2.6 illustrates the 180 UTM zones.

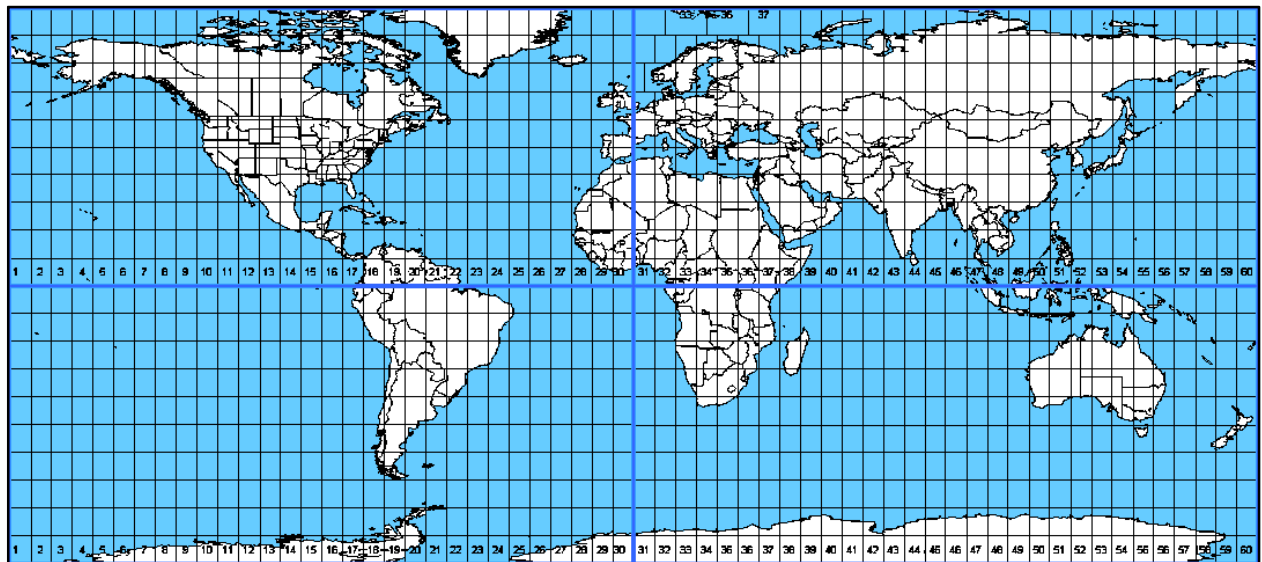


Figure 2.6: Universal Transverse Mercator projection (Resurgent Software, 2011)

2.3.2 3D GIS

ArcGIS 10 offers a 3D environment, which presents a three dimensional viewing description of the GIS data, as well as provide some analysis in 3D in ArcScene. Prior to ArcGIS 10, GIS was considered to only be able to offer a 2.5D environment. This was because it mostly offered a 3D environment (3D Analyst extension) as a viewing tool. The original analysis offered by 3D Analyst mostly focused on surface analysis (Tiede & Blaschke, 2005; Bratt & Booth, 2004).

With the release of ArcGIS 10, 3D Analyst added new analysis tools to the old version such as data editing techniques in 3D, where the user has the option of assigning a z-value to the data. This z-value is taken into consideration when analysis queries are run on the data. Existing geoprocessing tools such as the “Select by location” (which highlights a feature base on its location parameters) and “Line of sight” (which determines the sight path

between two objects) tools are enhanced to perform better in a 3D environment. One of the main benefits of the new 3D Analyst is using model builder to create models which can do analysis on a network dataset in 3D (ESRI, 2011). A network dataset will be defined clearly at a later stage in this chapter.

This analysis, although a good improvement from the previous version, still does not offer complete three dimensional analysis tools. There is a great need for a 3D GIS data model which offers tools to analyze, visualize and perform spatial queries on three dimensional data. One of the main inadequacies is the lack of three dimensional topological rules. Topological rules enhances the integrity of the data by defining the manner in which geographic elements may share spatial geometry. For instance telecommunications network switches (point features) may only be located indoors (inside a building polygon). However, present commercial GIS software only offer topology rules in a 2D environment, and do not take into account three dimensional height values (z-values). A satisfactory 3D topology would need to include primitive two dimensional geometry types, such as point, line and polygon, as well as primitive three dimensional shapes. Primitive three dimensional shapes include cubes, spheres and cylinders (Actur & Zeiler, 2004; Lee & Kwan, 2005; Ellul & Haklay, 2006).

Analysis in a 3D environment is however possible to a degree, when exploring outside the realm of ESRI software. According to Hijazi *et al.* (2010), by using alternative software such as Open Source, a designer can create his/her own 3D environment. Network analysis is possible in an Open Source environment, but developing such an environment requires extensive programming skills, whereas ESRI software is more user-friendly.

2.3.3 Feature representation in GIS

At the heart of GIS, the system aims to represent natural and/or man-made elements visually in order to create a better understanding of the features by sorting them into thematic layers. A thematic layer is a collection of common geographic elements such as parcel boundaries, road networks and point locations (Actur & Zeiler, 2004).

Geographic features are represented in a GIS in two ways. The first is vector data. Vector data represents the shapes of features precisely and compactly with a set of coordinates and attributes. Vector data consists of four data types. Points are features which are too small to be represented as lines or areas. Stop signs is a typical example of point data in a transportation GIS. A second data type, are lines. These are features which are too narrow to be represented as areas. Normally these are long, linear features such as streets or rivers. Polygons are elements which represent area features. Polygons are a series of segments, which encloses an area, such as parcel or municipal boundaries. The final vector data type is annotations, which are only valid in a geodatabase. These are descriptive labels which describe names and attributes and are stored as an independent layer (Zeiler, 1999).

Features can also be represented as raster data. A lot of geographic data is presented in grid form as a continuous surface, because cameras and imaging systems record data as grid cell values. Although a cell covers a certain predefined area size, which in its own may contain various features, the cell is only awarded a single value. A pixel is one cell of a grid, and has one value for the whole pixel. Normally the value of the pixel is determined by the most abundant feature in the particular pixel. A raster can represent a variety of different elements such as vegetative type, surface value or elevation (Zeiler, 1999).

2.3.4 Data storage methods

There are several ways to store geographical data in a geographic information system, such as coverages, shapefiles and geodatabases.

2.3.4.1 Coverages

The original method was to employ coverages. In earlier ArcGIS software versions, coverages were employed to store vector data. Non-spatial tables are not supported by coverages. Coverages can however be displayed, queried, analyzed and edited in ArcGIS. It employs feature classes, which is a collection of similar feature elements, representing the same geographic elements such as cables, switches or soil types (spatial tables). Each

feature class is a point, line, polygon or annotation. An annotation is a text which describes the feature. More than one feature class is normally needed to represent a feature in a coverage. For example a coverage which represents a polygon feature incorporates point as well as line feature classes to determine the polygon area. Polygon coverages also contain a label points feature class (it determines the location of labels on the feature), as well as a feature class containing tic points. Tic points do not represent any geographic data but represents the extent of the coverage on the real-world coordinates (Childs, 2001; Actur & Zeiler, 2004, ESRI, 2011).

2.3.4.2 Shapefile

The successor of the coverage is a storage type named shapefile, which also stores vector data. Shapefiles were used to store non-topological geometry and attributes of spatial features. The geometry of a feature is stored as a shape comprised of a set of vector coordinates. Shapefiles can support point, line and area features, and are often stand-alone files, each with its own coordinate system. Shapefiles do not support non-spatial tables, nor do they have any topological overhead and thus cannot employ measurements to ensure data integrity. This data storage type is normally utilized to create a short term geographical representation such as creating and exporting maps and is not suitable as a permanent data management tool (ESRI, 1998).

2.3.4.3 Geodatabase

The final data storage method in ArcGIS is the geodatabase. Pacurari (2002) defines a geodatabase as a repository, which contains spatial data such as vector data, raster data and non-spatial tables. The term geodatabase is short for geographic database, a relational database containing geographical information.

A logical data model is the thematic layout and feature description of a geodatabase, while a physical data model consists of the physical implementation of the data model. The geodatabase data model brings a physical data model closer to its logical data model and

contains the same objects as in the logical model. The geodatabase also lets the user implement object behaviour without the necessity for extra code being written. These behaviours are applied through domains, relationship classes, topologies and other functions provided by the geodatabase (Zeiler, 1999; APDM, 2004).

A geodatabase stores four types of data (Pacurari, 2002; Zeiler, 1999):

- Vector data: Vector data are features with discrete spatial boundaries.
- Raster data: Raster data divides space into uniform cells or pixels. A raster dataset has a two-dimensional matrix and stores a value for each cell such as a satellite image.
- Triangulated irregular networks (TINs): A series of irregular spaced points, each having a value which describes the surface at that point (e.g. elevation). A network of linked triangles is constructed between these points, and where adjacent triangles share two nodes and an edge (side), they are connected to form the surface.
- Addresses and locators: Addresses are used to find the location of features on a map. Locators can also contain information that allows the user to create new features for a location.

The advantages offered by a geodatabase can be summarized as:

- It offers a central storage and management point for all data.
- Data integrity is enhanced through subtypes, domains, validation rules and relationship classes.
- A geodatabase offers tools for the creation of feature linked annotations and geometric networks.
- It offers fast and efficient data entry
- Feature shapes are well defined in a geodatabase and this leads to the production of higher quality maps.
- It promotes multi-user editing sessions in an enterprise environment.
- A geodatabase easily integrates with other software (Zeiler, 1999; APDM, 2004; Murphy, 2004).

Due to the fact that a geodatabase improves the administration, access, management and integration of spatial data when compared to both a shapefile and a coverage, it was decided to implement a geodatabase as the data storage method for this study (Childs, 2001).

2.4 Geodatabase concepts

Section 2.4 provides a detailed description of a geodatabase layout. It describes the function of feature classes; subtypes; feature datasets; topologies; geometric networks; and network datasets. The section also illustrates the purpose of relationship classes; domains; and how tables and rasters are stored in a geodatabase. The section concludes by describing the function of a fishnet, a toolbox and ModelBuilder, and how each is stored.

2.4.1 Feature class

In a geodatabase, spatial features and their attributes are stored in a feature class. Zeiler (1999) describes a feature class as a set of features of the same kind and the same geometric shape. Feature attributes are properties which defines each feature. Every feature in the same feature class, share the same numeric, descriptive or textual fields (columns in the table). A feature class stored in a geodatabase may be divided into subtypes. Subtypes are when features share the same attributes and geometry types, but have different meanings in terms of the data model. An example is that a feature class containing roads has a numeric field called type, which has a range of 1-3. If the type value is 1, then the type of road is a highway. In the same way, when the values are 2 or 3, the road type is primary road or secondary road respectively. Subtypes also allow the designer to predefine a set of attributes for a certain subtype. Still considering the roads example, assume that the highway has a maximum car capacity of 2 000 cars per hour, while the primary and secondary roads have maximum car capacities of 1 000 and 500 cars per hour respectively. When creating subtypes the designer sets the maximum car capacity for each, to its own unique default value. When a new road is created, and its subtype is set to highway, the maximum car capacity automatically sets to 2 000 cars per hour (Actur & Zeiler, 2004; Pacurari, 2002; ESRI, 2011).

2.4.2 Feature dataset

Similar or related feature classes are grouped together in a feature dataset. A feature dataset is also a storage place for the relationship classes between features, the topology rules and the networks. All the related feature classes that interact with one another in a feature dataset must have the same coordinate system. This specifies the dataset's real-world location, which includes properties such as map projection, the datum and the allowable range of the coordinates (extent). The feature dataset helps to increase data integrity in the geodatabase, as well as enhances the management of the data. The ability to store a feature class inside a feature dataset is the fundamental advantage when implementing a geodatabase (Pacurari, 2002; Zeiler, 1999; ESRI, 1998; Actur & Zeiler, 2004).

2.4.3 Topologies

Topologies are sets of predefined rules in a GIS, to secure spatial integrity. Topology rules define the manner in which different features may share geometry. They manage a set of topology rules, and define valid spatial relationships between features. Topology rules may also define how a feature class interacts with itself spatially. These rules in essence define the adjacency, connectivity and containment between features. For example, land parcels may not overlap with one another. With the release of ArcGIS 10 in 2010, ArcGIS offers 32 different topology rules to choose from. These consisted of 10 polygon-, 6 point- and 15 line topology rules. ArcGIS 10 also offers one hybrid line- and polygon topological rule. Examples of some of the rules may include: voting districts that must be covered by countries, or lines that may not intersect. A topology, which may manage many rules, is stored in a feature dataset, and only applies to feature classes stored in the same feature dataset (ESRI, 2011; Actur & Zeiler, 2004; Ellul & Haklay, 2006; Lee & Kwan, 2005; ESRI, 2003).

Zlatanova *et al.* (2003) stated that in order to consistently maintain a high level of data integrity, topologies are of the utmost importance. The advantages of using topological structured objects in a GIS, was listed by Baars *et al.* (2004) as:

- Avoids redundant storage.
- Maintains consistency of data after editing.
- Supports certain visualization or query operations.
- It helps detect topology errors in the feature dataset.

Topologies offer the option to mark certain errors as exceptions. In an ideal world, topologies offer rules to perfectly manage spatial relationships between features. The real-world however may contain situations which are exceptions to the rule. The geodatabase is flexible enough to accommodate these exceptions. Violations of the topology rules are initially stored as errors. These errors can be marked as exceptions, which are then ignored (ESRI, 2011).

2.4.4 Geometric network

A geometric network is a connected network consisting of a set of selected edges, junctions and end points. The participating feature classes and each one's role in the network is defined by the geodatabase designer, and organized in a feature dataset. Zeiler (1999) defines a geometric network as an element which models linear systems and supports network-tracing and –solving functions. Geometric networks are mostly used to represent the connectivity of a utility networks, such as water or electrical networks. Geometric networks define the connectivity rules between features in a feature dataset. For example a 10 inch water main can only connect to an 8 inch main through a reducer.

A geometric network represents networks which flow in a singular direction, such as water in a river. All the lines of the network are represented as edges, while the points are represented as junctions (or end points at the end of the network) respectively. Polygons cannot take part in a geometric network. The network contains a source (origin of the network) and sinks (end points of the networks). By assigning a source and sink to the network, the user defines direction to the network and is able to perform analysis on the network (ESRI, 2011, Actur & Zeiler, 2004; Pacurari, 2002). Figure 2.7 illustrates a water distribution network. Water mains and service lines (edges) are connected to one another through a water junction.

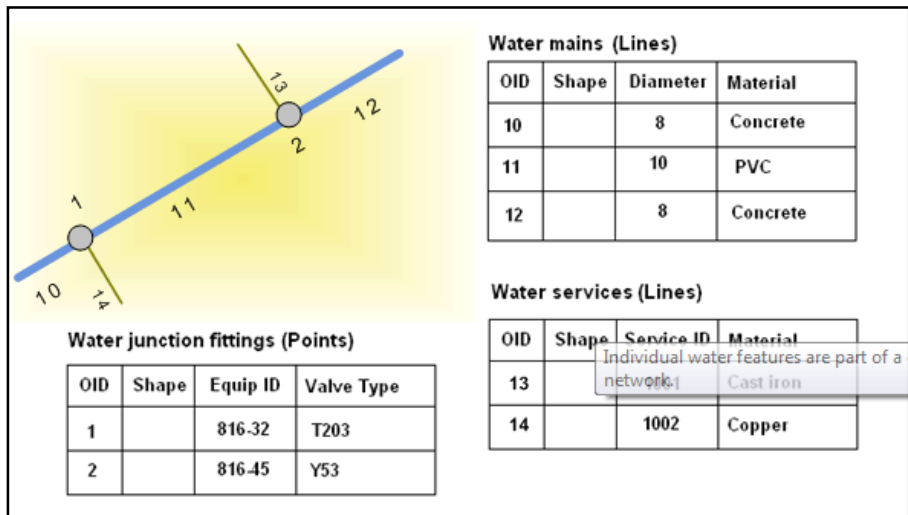


Figure 2.7: A geometric network for a water distribution network (ESRI, 2011)

The Utility Network Analyst toolbar in ArcMap is used to perform analysis on the network. The downstream tracing tool displays all the features downstream from a certain marked point, while the upstream tracing tool displays all the features located upstream. The “find the shortest path between two points” analyses the network and presents the path with the least resistance between selected points. Finally, the “find a common ancestor tool” displays a common root directory between features (ESRI, 2011).

2.4.5 Network dataset

Network datasets are more commonly used for networks which flow in both directions, such as transportation networks. A network dataset does not work in a hierarchical manner such as geometric networks. Just like a geometric network, the network dataset represents points and lines as junctions and edges respectively, but has no endpoints. A network dataset is an integrated network which moves in both directions along an edge, but may also contain certain restrictions (such as a one-way road). It also stores the datasets connectivity rules, directions and turn delays. For example a South African transportation network (left hand side of the road) may have a longer delay to turn right at an intersection, compared to when turning left. The network dataset saves this information, and uses it to do analysis such as finding the quickest route between two points. A network dataset can also be implemented for indoor use. One such example is a network dataset being employed to represent the hallways and staircases inside a building, in order to determine the quickest route between two points (ESRI, 2011; Actur & Zeiler, 2004; Crickard, 2010).

Analysis on the network dataset can determine the service area from a certain point. A network service area is a region which covers all the edges which are within certain distance from a feature. A route layer is created when route analysis has been performed on the network. The route layer stores and displays all the inputs, parameters and results of the route. The closest facility tool determines the closest feature to an incident on a network, while considering the cost parameters of each edge. Network dataset analysis can be performed in ArcMap by utilizing the Network Analyst toolbar, or by using the geoprocessing tools provided by ArcGIS. Either the ArcEditor- or the ArcInfo version of ESRI's ArcGIS product is required in order to create and edit network datasets and geometric networks. The Network Analyst extension is also required in order to perform analysis on a network as well as gain access to the network toolbars (ESRI, 2011).

2.4.6 Domain

A domain is a set of possible values for a field. A domain is saved for the whole of the geodatabase, and can be used by any field for which it is appropriate. There are two types of domains: a coded domain and a range domain. A coded domain has a list of integer value, each of which represents a different description. For example, assume in a telecommunications network, wire can only be a fibre optic wire, a UTP (unshielded twisted pair) or a telephone wire. By creating a domain called wire, with value: 0 = Fibre optic; 1 = UTP and 2 = Telephone, the designer makes it impossible for another editor to define the wire as something else, therefore eliminating human error. In the same way, a range value is only valid if it is within its range. An example of this domain would be to create a domain for the pH-balance field. It is only valid for a range between 0 and 14, any other value is impossible. Seeing as the domain is saved for the geodatabase, any table or a feature class with a similar field (pH-balance) in another feature dataset, may also utilize the domain (ESRI, 2011; Actur & Zeiler, 2004).

2.4.7 Relationship class

A relationship is an association between two features or objects. Relationships are managed and organized in relationship classes. A relationship class describes a set of relationships

between feature types. Relationship classes depict how two tables (spatial or non-spatial tables) relate to one another, and make use of primary and foreign keys to relate to each other. A primary key is a unique identifier of a table (origin table), such as a list of citizens having an Identification number. Each individual entry (row) of the table has a unique ID, normally the ObjectID field. A foreign key in another table (destination table) represents the same ID number. By creating a relationship class between the two tables using the primary and foreign keys, the information in the origin table, for that specific primary key, is automatically linked to the entry in the destination table. When the identifier tool in ArcMap clicks on a feature, the feature's attributes are visible, as well as the attributes of the related feature. Relationship classes reduce redundant fields in the geodatabase (ESRI, 2011; Actur & Zeiler, 2004).

There are different types of relationship classes depending on cardinality. A relationship's cardinality represents the number of features in the origin table that can relate to a number of destination table features. There are three types of cardinalities:

- One-to-one relationship: This is where a feature may only be linked to one other feature. An example is that a husband may only have one wife.
- One-to-many relationship: In this relationship, an origin table feature may be linked to many destination table features, but a destination table feature may only be linked to one origin table feature. For example an owner may have many cars, but a car may only have one owner.
- Many-to-many relationship: In this relationship an origin table feature may be linked to many destination table features, and vice versa. An example of this relationship class is when a building may have many owners, and an owner may own many buildings. In order for this to be possible an intermediate table needs to be created with two fields, each containing the primary keys of tables. These are both foreign keys to the respective tables, so the unique identifiers may repeat. This allows both tables to have many links to one another (ESRI, 2011).

Relationship classes can be either simple or composite. In a simple relationship class, related objects can exist independently of each other. When an origin object is deleted, the related destination object's value is set to null. In a composite relationship class a destination object can't exist independently of origin objects. When an origin feature is deleted, a destination object is also deleted in a process called cascade delete (ESRI, 2011).

2.4.8 Tables

Tables are stand-alone elements in a geodatabase. A table consists of rows and columns. Each row in a table represents a table feature and is called a record. Each column represents an attribute value and is called a field in GIS. Tables can also be joined to one another by using a shared identifier field to connect the tables. Relating tables defines a relationship between two tables, but doesn't append the attributes, also utilizing a common field. The data can be accessed when necessary. Joins and relates lack some of the advantages offered by a relationship classes such as composite objects, referential integrity, messaging, attributes and relationship rules (ESRI, 2011).

2.4.9 Rasters organization

Raster data can be stored in four ways in a geodatabase, such as:

- A raster dataset is a single picture or a seamless image covering a spatially continuous area.
- In a mosaic dataset, a collection of raster datasets is saved as a catalog. This allows the user to store, view, run queries on and manage collections of raster datasets.
- A raster catalog is a collection of raster datasets displayed as a single layer. Each one may have different coordinate systems or data types.
- The final method saves a raster as part of a feature's attributes, such as a column of a feature class.

The data model in this study will contain one raster, and therefore it was decided to implement a raster dataset (ESRI, 2011).

2.4.10 Toolbox and geoprocessing tools

The geodatabase offers the ability to create and store geoprocessing tools. Geoprocessing is a concept which facilitates the automation of GIS tasks. A large number of GIS implementations involve repetition, which creates the need for the automation of GIS

methods. Geoprocessing supports automation by offering a collection of tools and techniques to combine a series of tools in a sequence by implementing models and scripts. One such geoprocessing tool is the *Create Fishnet* tool. The create fishnet tool creates a feature class consisting of a net of rectangular cells (ESRI, 2011).

These tools and sets of tools can be stored in a toolbox, which in turn can be saved in the geodatabase. ModelBuilder is an application which creates, edits and manages models. Models are workflows that create sequences of geoprocessing tools. The output of one process is used as input for the next. ModelBuilder starts off as a blank canvas which gets populated by geoprocesses by dragging them from a toolbox. ModelBuilder validates the sequence, run the geoprocesses and finally alters the map according to the final output (ESRI, 2011).

2.4.11 Three types of geodatabases

ArcGIS present three types of geodatabases, the personal geodatabase, the file geodatabase and the ArcSDE geodatabase (see Table 2.4).

The personal geodatabase was introduced since the initial ArcGIS release (version 8.0). The data in a personal geodatabase is stored and managed using Microsoft Access (.mdb file). This geodatabase is ideal for a single user, using small datasets. The maximum size for the entire geodatabase is 2 gigabyte, although the effectiveness of the database decreases after the size reaches between 250 and 500 megabytes. This type of geodatabase is only supported by Microsoft Windows. The personal geodatabase supports all normal geodatabase information types such as topologies, raster catalogs and network datasets. The personal geodatabase does not support multi-users and versioned workflows. The personal geodatabase supports attribute manipulation and many users make use of a personal geodatabase due to its great string handling in Microsoft Access (Actur & Zeiler, 2004; ESRI, 2010b).

Characteristics	Arc SDE	File geodatabase	Personal geodatabase
Description	Various types of GIS datasets, stored in a DBMS	Various types of GIS datasets, stored in a file system folder	Geodatabase stored and managed in Microsoft Access data files
Number of users	Multi-users. Many editors and many users	Single user and small workgroups. Some users and one editor per dataset	Single user and small workgroups. Some readers and one editor
Storage format	Oracle, SQL Server, IBM DB2, IBM Informix	Datasets are files within a geodatabase file	Microsoft Access (.mdb)
Size limit	Up to DBMS limits	1 Terabyte per dataset	2 Gigabyte per geodatabase
Operating system	Any platform	Cross-platform	Windows

Table 2.4: Differences between an ArcSDE geodatabase, a file geodatabase and a personal geodatabase (ESRI, 2008b)

In the file geodatabase, each dataset is stored as a separate file and each geodatabase is accumulated in a file folder. This database provides fast performance and can handle very large file sizes. Each dataset can reach a maximum of 1 terabyte. The file geodatabase far exceeds the personal geodatabase in terms of storage capacity and performance. The file geodatabase also supports all regular geodatabase information types except multi-user editing and other version-based workflows. The goals of a file geodatabase are: to offer an available geodatabase solution to a broad variety of users; to provide a geodatabase which operates across a different operating systems; and to provide superb performance while handling large files (well over 300 million features and datasets, scaling beyond 500 gigabyte each). A file geodatabase uses about a third of the feature geometry required by shapefiles and personal geodatabases. In order to reduce the required storage space, it offers compression of vector data into a read-only format (ESRI, 2008b; ESRI, 2010b).

An ArcSDE geodatabase manages its data within a DBMS architecture such as DB2, Oracle, SQL Server and Informix. ArcSDE can handle extremely large continuous datasets

and offers a multi-user editing environment as well as version-based workflows such as archiving and geodatabase replication. ArcSDE supports many simultaneous users and inherits the benefits of a DBMS including reliability, security and backup (ESRI, 2008b; ESRI, 2010b).

There are three levels for accessing ArcSDE in ArcGIS : the personal SDE, the workgroup SDE and the enterprise SDE. The Personal ArcSDE provides the ability to fully administer and manage the ArcSDE geodatabase, which provides full capabilities for a small amount of users and one editor. The Personal ArcSDE geodatabase uses SQL Server Express. The Workgroup ArcSDE can also access, fully administer and manage the ArcSDE database, and can handle up to ten simultaneous Windows desktop users and editors. The Workgroup ArcSDE also uses SQL Server express. The Enterprise ArcSDE database size is only limited by the DBMS it utilizes (DB2, Oracle, SQL Server or Informix) and can handle an unlimited amount of users (ESRI, 2008b).

An ArcSDE geodatabase is more suitable for large enterprises which have the need for multi-user access on a large scale, and was therefore not considered for this pilot project. It was decided to implement a file geodatabase due to the fact that it offers more benefits than a personal geodatabase in terms of performance, size and cross-platform interoperability (ESRI, 2008b; ESRI, 2010b).

2.4.12 Geodatabase design

The success of a GIS project has a very strong correlation with the design of the geodatabase. In order to successfully achieve the project goals, the geodatabase has to be designed according to the characteristics of the features as well as the end-product (Zeiler, 1999). With the intention to achieve the goals set by this project, it was decided to compare three geodatabase design methods with one another. The method that offers the most benefits was applied to this study.

When Buliung & Kanaroglou (2004) completed a study on developing a spatial database for micro level urban processes, they implemented a geodatabase design technique which consisted of four phases (Figure 2.8). The first step entailed a definition of the real-world aspects as well as the final objectives of the database. The second phase is called the conceptual modelling. This phase determines the features that will be represented in the database. The penultimate phase is the database design phase. This phase divides the features into classes, determines relationships and defines the database functions. Finally the fourth phase implements the database by physically creating a database.

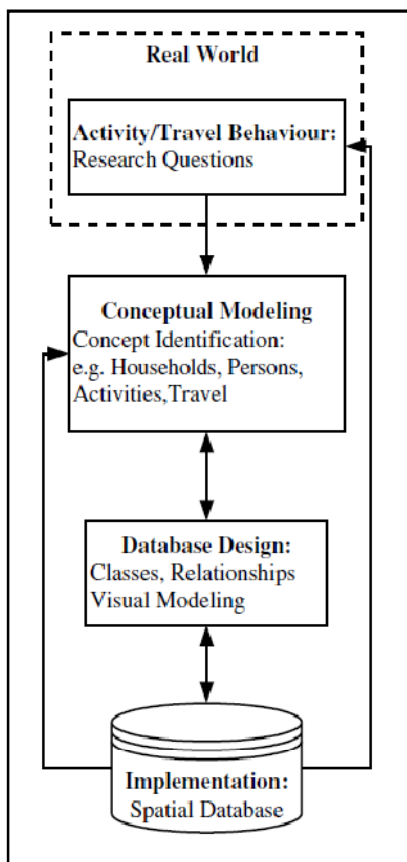


Figure 2.8: Four phases of database design (Buliung & Kanaroglou, 2004)

In the method described by Zeiler (1999), there are five design steps defined. Step one is to create an understanding of the objectives of the project as well as determine the data needed to achieve these objectives. Step one also divides the data into logical layers. Step two defines every object which will be represented, as well as the relationships they have with one another. The geographic representation of each object is decided in the third step. The fourth step provides the features with attributes. The final step organizes the geodatabase structure. In this step the object classes; relationship classes; feature classes;

geometric networks and topologies are organized in the feature datasets. In turn the feature datasets, raster datasets and tables are organized in the geodatabase. The 10 step design method described by Actur & Zeiler (2004), extends the method described by Zeiler (1999) as seen in Table 2.5.

Step	Description
1	Identify the final product
2	Identify the key thematic layers
3	Define the scale ranges and spatial representation of each layer
4	Arrange similar or relevant data into datasets
5	Define the tabular database structure
6	Define the spatial properties of the dataset
7	Propose a geodatabase design
8	Implement, prototype, review and refine the design
9	Design workflows for designing and maintaining each layer
10	Document the design

Table 2.5: The 10 steps to designing a geodatabase (Actur & Zeiler, 2004)

The first step is to identify the final information product that will be produced by the GIS. In this step the primary objective of the GIS must be defined. The objectives may include inventory maps, analytical models, data flows or Web access. The second step is to identify the key thematic layers, based on the information requirements. This step identifies which features will be represented, and at what scale and which symbology will be applied for each. Step 3 in the design is the specification of the scale range and spatial representation of each layer. In this step the feature representation is defined as a point, line or polygon.

The fourth step in designing a geodatabase requires the feature classes to be grouped into datasets. Different features are modeled with feature datasets, feature classes, relationship

classes and domains. Continuous data is modeled with raster datasets, while measurement data is modeled with survey datasets. The next step is to define the tabular database structure and behaviour for attributes. This step defines attribute fields, domains, subtypes and relationships. Step six defines the spatial properties of the datasets. Networks are used to model connected features and topologies are used to maintain spatial data integrity and spatial geometry. The spatial reference is also set for the dataset. Step seven designs the geodatabase and puts it to paper and structural elements are applied.

Step eight is where a geodatabase is built and the data is loaded. This step incorporates the implementation, prototyping, review and refining of the design. Step nine designs work flows to build and maintain each layer as well as provide the metadata for the data model. The final step is to document the geodatabase design. Drawings, layer diagrams, schema diagrams and reports can be used to communicate and describe the data model.

Actur & Zeiler (2004) describes a geodatabase in detail, which provides much more information than the summarized version provided by the Zeiler (1999) model. The 10 steps also provide a sense of chronological sequence in which a geodatabase design should be undertaken (Actur & Zeiler, 2004).

If the methods described by Buliung & Kanaroglou (2004) and Actur & Zeiler (2004) are combined, it is possible to divide the 10 steps into three phases. Steps one through three describe the conceptual phase, which creates a basic understanding of the objectives of the end product. Steps four through seven describe the logical design phase. In this phase the features as well as their spatial representation, relationships, attributes, domains and topologies are defined. The final phase is the physical design phase (steps eight to 10) and consists of physically creating, prototyping, reviewing and documenting the geodatabase. It was decided to implement a hybrid combination between the methods proposed by Buliung & Kanaroglou (2004) and Actur & Zeiler (2004) for this study.

2.5 Case studies

Acquiring published articles on GIS geodatabase data models was a tedious journey. Three of the four following case studies were presented at GIS conferences. The tendency seems to be for GIS professionals to present their work at user conferences rather than publish an article of their work. Only one of the case studies was published as a dissertation. The aim of this section is to understand how previous studies have implemented GIS in telecommunications or other network related sectors and how these methods and techniques can be implemented in this study. All four case studies implemented GIS with different data models and different objectives.

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

The aim of the project was to construct a GIS data model for navigation through urban environments. The study area was the north campus of the University of Buffalo. Mandloi (2007) developed a GIS data model to represent a multi-modal transportation network, for inside as well as outside of buildings. A single modal network represents a single method of transportation such as driving in a car, whereas a multi-modal network depicts a series of intertwined transportation networks. An example of the latter is a representation of a study area's combination of the bus routes; train routes; cycling lanes; walkways; and the points of transfer between the methods.

A big concern when developing the project was depicting multiple level buildings. It was decided that the transportation network inside buildings would be represented in three dimensions, while the outside transportation network was represented in 2D. By doing this the designer aimed to create a seamless integration between inside and outside.

The final objective of the study was to develop an application which uses the data model presented in the study and integrates geo-spatial information such as aerial photographs, CAD drawings and paper maps in order to provide a shortest path analysis in 2D and 3D.

The construction of the project was divided into two sections: modeling movement inside of a building (3D) and modeling movement outside a building (2D). A multiple level building can essentially be seen as several horizontal blocks on top of one another, connected vertically via elevators and staircases. A building can be divided into floors, and floors are sub-divided into rooms and corridors. Each floor has its own horizontal network and is connected to each other vertically. One problem encountered during the development of indoor transformation was that a corridor plays a major role in the movement. A corridor is however a polygon, and only lines and points are allowed to take part in a network. As a solution to this headache, the designer created a medial axis of the corridor polygon, to be used as a line in the transportation network.

Another problem was the representation of multiple floors in a geodatabase. Mandloi (2007) considered three options. The first option was to represent each floor of each building in the study area in an individual polygon feature class. Even in a very small study area, the amount of feature classes would be immense and data management confusing.

The second alternative was to create one feature class for each floor, covering the whole study area. The number of feature classes would be the same as the number of floors of the highest building. This method would cut the number of feature classes in the database by a substantial amount. The problem with this method is that the feature classes representing the lower levels would be highly populated, while the feature classes of the higher floors would be sparsely populated. For example, only some buildings have a tenth floor, while all buildings have a ground floor.

The final option was to represent the total number of floors in all the buildings in one feature class. The different floor values of the rooms can be maintained by implementing subtypes. When drawing a room polygon, the z-value of its corners can be set so as to coincide with its floor value, thus displaying on the right height in ArcScene. This alternative minimizes the size of the database exponentially. The only drawback is that when a very tall building is created, the developer must stay extremely focused when entering the z-values in order to avoid mistakes.

It was also important for Mandloi (2007) to calculate and model the time used to move in the building statistically. The time used to take an elevator as opposed to taking the stairs was crucial to develop the impedances of the network dataset at a later stage. The impedances define the delay time of a route, which helps to determine the shortest route.

Modeling movement outside the building was done in two dimensions. Making sure that the different line and point types were spatially connected was of the utmost importance. If the different networks weren't connected spatially to one another, they would not be connected in the network dataset. The outside multi-modal network consisted of the street network, predefined bus routes, walkways, and the transfer points between the specific modes of transportation. Due to the fact that the outside transportation model is not very relevant to the data model presented in this dissertation, it will not be discussed in more detail.

The GIS data model, together with its background layers were created using ESRI software, ArcGIS v.9.1. The coordinate system that was implemented was UTM Zone 18N (projected coordinate system), with a NAD1983 datum. CAD-drawings of the building layout, which were not geo-referenced were implemented. The CAD files were referenced to the coordinate system using a world file. A world file is a text file that provides the X and Y coordinates for two specified points on a CAD file, as well as their respective X and Y coordinates in the projection coordinates. Geo-referencing CAD files and creating world files will be discussed in more detail in Chapter 3. The referenced data that was used for the project was one foot digital ortho-photos of the study area.

The floors feature class was digitized from the CAD drawings. Each room in the floors feature class had a floor_number value depicting the floor of the room and a z-value, representing the height of the room. Floors were only used as a visualization tool and did not play a part in the network. The total number of subtypes of the floor feature class was equal to the maximum number of floors of the highest building in the study area. The data model also contained feature classes representing exit points; floor turns; walkways; streets; campus bus routes; campus bus stops; parking lot entrances; sidewalk to bus route transfer points; and street to walkway transfer points.

The final step of the project was to create and test a network dataset for a best path finding application. A network dataset was created from all the feature classes participating in the overall network. The network dataset was utilized in the development of a 3D path finding application, named "Network3D". Network3D was created as a toolbar in ArcMap using ArcObjects and Visual Basic 6 programming languages. The application used Network Analyst to perform network analysis such as find best route. The application also used 3D Analyst to display the best route visually.

The path finding application can be applied to determine different types of queries. When tested it was found that the Network3D could determine the least effort route inside a building; determine an emergency route inside a building; determine the least effort route between two buildings; and determine the "all-inside" least effort route between two buildings (enclosed in a walkway).

The pilot data model that will be presented in Chapter 3 and 4, is similar to the one described by Mandloi (2007). This study will also represent infrastructure inside multiple level buildings. The IT utility infrastructure indoors will be represented in 3D, while the outdoor data will be represented in 2D, in order to create a seamless integration between indoors and outdoors.

The pilot project encountered a similar problem of representing the floors in a feature class. Similarly all three approaches, as suggested by Mandloi (2007), were considered and a decision was made to represent all the floors in one singular feature class, with z-values representing floor height. Another similarity is that the building layouts were only available in CAD-drawings that were not geo-referenced. The final resemblance between the pilot data model and the case study was that a network dataset was also created in order to develop a 3D routing application. The pilot data model however implemented ModelBuilder instead of ArcObjects and Visual Basic 6 to create the application.

2.5.2 Managing a fiber optic network with GIS – Montana-Dakota Utilities Company (Petersen & Taber, 2007)

The Montana-Dakota Utilities Company was founded in 1924, and acquired small electric and gas companies in North-Dakota and Montana. At the moment the company serves around 500 000 energy customers. The company started using fiber optic networks in 1998, but only for internal use. In 2000 – 2001 it underwent a massive network expansion and started to provide dark fiber (fiber optic infrastructure which is installed for potential future usage) to customers. The company’s expansion continued up to 2007, and now serves hospitals; telecommunication companies; the police and city government and has to manage over 50 miles of fiber optic cables in four communities.

In 1998 the company applied GIS in order to manage their electric and gas infrastructure. In 2006 the company again decided to apply GIS for the management and maintenance of their fiber optic systems. Fiber optic systems have less graphic data than electric or gas infrastructure, but have very complex relational data.

The system solution required a number of characteristics. Firstly the new system had to extend the current enterprise GIS of the company, and data had to be made accessible to the whole company. System analysis needed to be improved, especially tracing tools and reporting tools. The system needed to model the core features of the fiber optic network as well as the switch points.

Network feature	Non network feature	Structures
Fiber cable	Fiber slack loop	Fiber structures
Switch point	Demarcation point	Poles
Splice point	Proposed fiber cable	
End device point		

Table 2.6: A summary of the Montana-Dakota Utilities Company fiber optic network (Petersen & Taber, 2007).

Table 2.6 is a summary of the GIS. It shows that although there are 10 features represented in the GIS, only the fiber cables, patch points, splice points and end device points make out part of the network. Thus tracing analysis can only be performed on these features. The data model did also contain non network features such as demarcation points and proposed fiber cable layouts, as well as structure features such as poles and fiber structures.

In conclusion, the GIS developers made use of the existing enterprise GIS and expanded it to accommodate the fiber optic network, which also allowed company wide access to the data. Company-wide access to the network data improves the one-call process and management can find the damaged part of the network easier, saving a lot of time and money. The implemented data model also improved the system documentation and allowed system analysis.

This case study describes the most important features that must be present in a good telecommunications data model, such as the different types of cables, the intersections of the network (switches, hubs and servers) as well as the end points of the network (ports). This case study uses a geometric network to illustrate its telecommunications network, but since the aim of this study is to represent and perform analysis on a telecommunications network in a multiple level environment in 3D, a network dataset will be implemented instead. According to Van Maren (2011), ArcScene can perform network analyses on a network dataset but not on a geometric network. Therefore a geometric network could not be implemented in this study.

2.5.3 Optimization of cable path in a WAN using GIS (Kumar *et al.*, 2004)

This case study describes the project in which the Birla Institute of Technology and Science, located in Pilani, Rajasthan, India aimed to provide approximately 3500 students and 600 staff's quarters with a wide area network (WAN). The project aim was to implement GIS to find an optimum path for both the fiber and copper cables and hence reduce the cost. Especially the usage of network analyst and the shortest path analysis tools was put into practice. The shortest path analysis finds the path with the minimum cumulative impedance between nodes in a network.

The main idea of the network is to have a master switch as the primary source. From the master switch optical fibers are laid outdoors to connect with distribution switches within the buildings. Each building may only have one distribution switch. From the distribution switch, copper cables are connected to each office or room. There should be only one wire catering to each room from the distribution switch. For example if there are 20 offices, there has to be 20 copper cables from the switch.

The methodology was initiated by obtaining ortho-images of the hostels. The interior view of the buildings were created using AutoCAD. The CAD drawings were drawn into ArcView, georeferenced and digitized in order to align with the image data and create shapefiles for the different themes.

The optimum path for the outdoor fibers and the indoor copper cables were done separately. Two “plans” were developed for the possible layout of the network, and analysis performed on each. In the same way, possible potential layout plans were compared for the copper cables. A network for the copper cables inside the buildings was created with certain constraints in mind: each building may only have one distribution box, the length of the wire is not allowed to exceed 90 meters, and there is only one separate wire catering allowed per room from the access switch. The network shows the distribution switch as the source of all paths. The distribution switch is also digitized (see Figure 2.9).

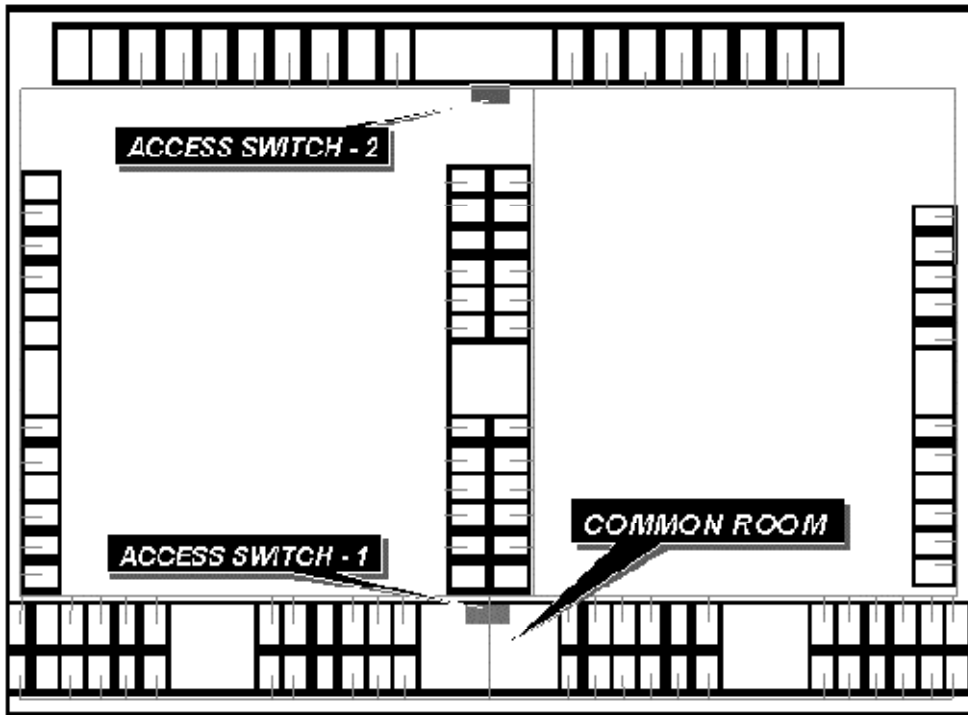


Figure 2.9: The location of the network cables and access switches in the hostel (Kumar et al., 2004)

The optimum path for the outdoor fibers and the indoor copper cables were done separately. Two “plans” were developed for the possible layout of the network, and analysis performed on each. In the same way, possible potential layout plans were compared for the copper cables. A network for the copper cables inside the buildings was created with certain constraints in mind: each building may only have one distribution box, the length of the wire is not allowed to exceed 90 meters, and there is only one separate wire catering allowed per room from the access switch. The network shows the distribution switch as the source of all paths. The distribution switch is also digitized (see Figure 2.9).

The aim of this case study is to show how CAD data for building interior can be incorporated into a GIS by georeferencing the CAD files and using it to digitize new building interior feature classes. The case study also emphasizes some of the limitations encountered when working with cable networks, which include that the length of the cable may not exceed 90 meters and that there is a separate wire from the switch to each end point. The case study also highlights the importance of creating the right relationship classes between features. For example a distribution switch may be connected to many cables, but a copper cable (subtype) can only be connected to one distribution switch. However a fiber optic cable

(subtype) may be connected to two distribution switches, thus this is a many-to-many relationship. A network port may also be only connected to one cable, and vice versa (one-to-one relationship). This case study, as described by Kumar *et al.* (2004), helped to define the required relationship classes among different features that would be implemented in this study.

2.5.4 Building and technology passport of Masaryk University (Glos, 2008)

Masaryk University is the second largest university in the Czech Republic. The university houses 9 faculties, 32 000 students and 4 000 employees. The campus consists of 200 buildings, which contains 17 000 rooms and covers a total of 355 000m². The majority of the university's telecommunications consists of data and voice networks. There are about 50 000 cables, 10 000 patch panels, 5 000 devices (excluding personal computers) and 20 000 data ports, which make the bulk of the network.

The aim of the project was to build a GIS for the campus buildings and their technologies. The goals of the project consisted of gathering, managing and providing data about buildings, its rooms and the technologies inside of the buildings. Other purposes of the GIS include: increasing buildings operations and administrations; enhancing the efficiency of building usage; and decrease the operational costs of the building.

The "passports" referred to in the title describes GIS systems created for the buildings and the technologies and were constructed separately. Both the building- and technology passports created their own data models. The building passport's data model included features such as: localities, buildings, floors, rooms, engineering constructions and the horizontal projection of the feature. The attributes in the building model include the feature properties (room number, room size) and height (for the 3D model). The attributes also included a positional code. The positional code is a coding system which allocates a unique number to each room (similar to a primary key). This code can also be used to detect on which floor, in which building and in what locality a room is situated. For example if: BNA is the locality number, BNA01 would be building 1 in locality "BNA". In the same way every

floor in the building has a unique extension code (e.g. N01), as well as every room in the building (e.g. 210b). This room example's unique code would thus be: BNA01N01210b.

Relationships in a data model are very important, in order to maintain data integrity when new data is entered. Some of the relationships used in this data model included: many buildings to one locality; many floors to one building and many rooms to one floor. The source data for both passports were derived from CAD (.dwg and .dgn) files, which were assigned a coordinate system and georeferenced to align with the rest of the spatial data, before it was used to digitize the feature classes.

Technologies can be situated in localities (outdoor switchboards), buildings (energy meters), floors (air units) and rooms (heaters), and therefore each technology has its own positional code. Technologies can also be a variety of systems and devices and were allocated a unique technology code, such as: systems (alarm system) – CA01; aggregates (alarm system for the 1st floor) – CA01EY001; and devices (alarm panel) – CA01EY001B01. Attributes in the technology data model include the feature name, its description, the type, date of installation and the date of last revision. The attributes also include the positional and technological codes.

The final result for developing the system was that the buildings and their technologies could be monitored and operated in a much more organized and cost-efficient manner. The system also presented the opportunity to develop a web-application version of the system which is accessible to employees and students.

The importance of implementing a positional code when developing indoor GIS and especially when working with multi-storey buildings is evident from this case study. The unique positional code extension for every locality, building, floor and room works as a foreign key, and the collection of positional extensions together form a unique primary key for each room. This study implemented a similar unique ID, as described by Glos (2008). Each case study described above differ from one another, but each one in its own respect highlights a method, technique or design template which can be incorporated into this study.

2.6 Conclusion

Chapter 2 discussed some of the basic network topology concepts as well as comparing GIS to alternative information management tools. Relevant GIS concepts were described, specifically including the different ways information can be represented in a GIS such as vector data, raster data and tables. Chapter 2 also compared three geodatabase design methods, and chose to implement a combination of two methods. The chapter concluded with four individual case studies describing techniques of how GIS have been implemented in order to accurately represent a telecommunications network.

The following chapter of this dissertation, Chapter 3: Methodology and Design, will describe the methods for the design of the GIS with the main focus on the geodatabase. By incorporating the ten steps defined by the combined design method, Chapter 3 explains the chronological steps of designing the geodatabase.