Chapter 3.

Designing the geodatabase

The conceptual design

Creating a well developed GIS system largely depends on the quality and accuracy of the spatial geodatabase provided for analyses (Mandloi, 2007). It is therefore extremely important to spend considerable time on database design in order to ensure successful implementation of the GIS system (Taggart & Ridland, 2000). As explained by Arctur & Zeiler (2004), the design of a GIS is composed of different sets of thematic layers representing common features. As indicated in Chapter 2, these thematic layers consist of features such as raster datasets and vector data comprised of points, lines and polygons. The accuracy of these data structures is of the utmost importance in order to provide an accurate representation of the real world data inside the GIS system.

The approach in designing a seamless geodatabase forms the core of this chapter and continues in Chapter 4. The focus of this chapter is mainly on the logical and conceptual design of the geodatabase. The following section, "database design", provides a brief outline with regard to general database designs. The basic database design concepts are indicated in order to provide a sufficient medium of comparison with designing a geodatabase.

3.1. Database design

Successful database design is based on a database "blueprint" such as the design drawings required for building a house. The designing phase of the database provides a structure of the information as well as the relationships between the data items that will go inside the database (Ul Haq, 2009).

According to Ul Haq (2009), four steps towards the design of the database are presented as:

1. *The requirement analysis*. This is the stage where all the requirements for the database are gathered and the techniques include:
   - selecting existing documentation, forms and other databases;

- research and site visits;
- observation of the work environment;
- questionnaires;
- prototyping - build a small model of the user's requirement to verify beforehand; and
- Joint Requirements Planning (JRP) - group meetings are conducted to analyze existing problems.

2. *Entity Relationship Diagram (ERD)* as represented in Figure 3.1. The process of expressing the database design graphically by means of simple and understandable symbols and diagrams. It provides a basic summary of the data gathered in the *requirement analysis.*
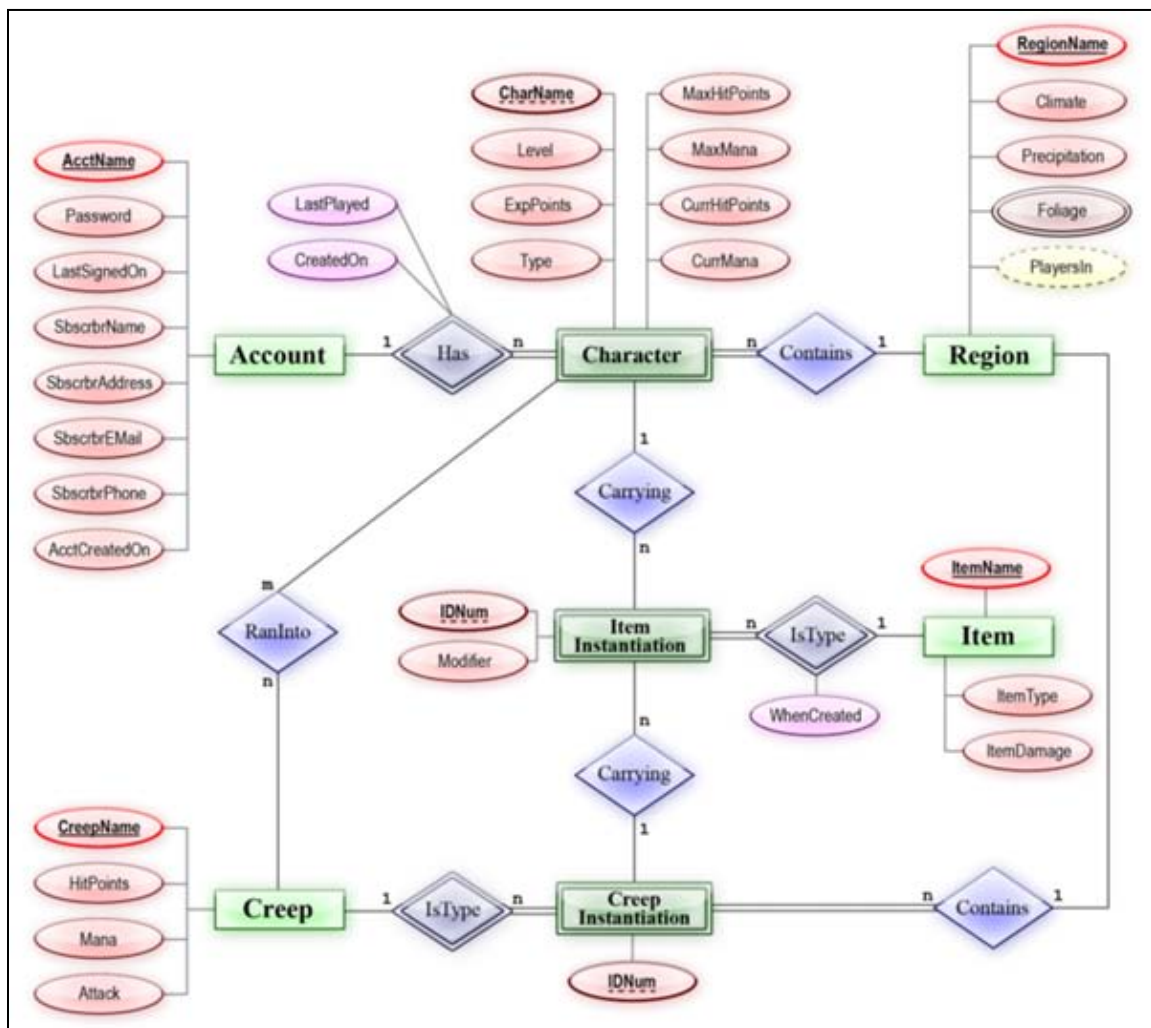

Figure 3.1. A sample Entity Relationship Diagram (Kaufmann, 1999)

3. *Relational model.* In this section all the relationships represented in the ERD are converted into tables in order to support the effectiveness of the implementation thereof. The data are converted in order to provide the possibility of implementing subclasses and relationships. In other words, this is the process of creating fields for the tables in the database with primary and foreign keys.

4. *Normalization.* In this section some dependencies are eliminated. Dependencies based on part of the primary key (partial dependencies) and dependencies based on an attribute that is not part of the primary key (transitive dependence).

Kaufmann (1999) provided his representation of the database design in the following steps:

- requirements analysis;
- logical design;
    - ER modelling;
    - view integration;
    - transformation of the ER model to SQL tables; and
    - normalization of tables;
- physical design; and
- database implementation, monitoring, and modification.

According to Mandloi (2007), the database schema consists of object classes, associations, and attributes and serves as the blueprint of the database. According to Booch, Rumbaughand & Jacobson (1999), the key phases in designing the database schema is as follows:

- identify the classes in the model;
- create class diagrams to describe the classes;
- expand the structural details of the classes. This includes specification of attributes and associations;
- check for common patterns that complicate physical database design (cyclic associations, one to one associations);
- consider adding behaviour to the object classes where needed; and
- construct a physical design (database) from the logical design (data model).

Heywood, Cornelius & Carver (2006) stated from their point of view that the database creation consists out of the following table as indicated by Oxborrow (1989) and Reeve (1996):

Oxborrow (1989)	Reeve (1996)
Data investigation	Needs analysis
Data modelling	Logical design
Database design	Physical design testing
Database implementation	Implementation
Database monitoring	Maintenance

Table 3.1. Steps in database creation (Oxborrow, 1989 and Reeve, 1996)

Referring to the abovementioned database design steps, a comparison could be drawn between all of the aforementioned.

It is compulsory to start with a requirement/needs analysis in order to analyze the type, quantity and qualities of the data that would be required for effective database design. In this stage the conceptual design of the proposed database is established.

The second phase is the logical design phase, also known as the data modelling stage, where a conceptual model of data is created. This is executed by looking at the relationships between entities, as well as characters of attributes and entities. It is also the stage where an entity relationship model could be created in order to provide a visual graphic representation of the data that would be used inside the database.

The third phase of the database design is the physical design also known as the database design that follows directly on the logical design of the database. This phase includes the creation of tables and fields that originate from the *entity relationship diagram (ERD)*. A design for the database management system is created and includes the creation of field names, field types and field structures.

The fourth phase consists of the database implementation, monitoring and modification. Attribute data are created for the database that is then monitored, fine tuned, modified and updated on a regular basis.

3.1.1. The ten geodatabase design steps (provided by Arctur & Zeiler, 2004)

With reference to Table 3.2 a holistic approach towards the creation of a geodatabase is provided. It resembles a complete and accurate way of creating a seamless geodatabase

especially suited for the study at hand. In comparison with the previous mentioned database design steps, the steps provided by Arctur & Zeiler (2004) in Table 3.2 offers a more detailed and complete description of the database design. This is especially when the focus is placed on geographical data. Referring to the database concepts presented in section 3.1, three main sections are present at each database design. These are:

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

It is also necessary to mention the fact that the database design provided by Arctur & Zeiler (2004) are specially designed for spatial data and are therefore also much more suitable for the geodatabase design than the other, which are designed for commercial databases.

Conceptual design	1. The information products that will be produced with the GIS need to be identified
	2. Based on the information requirements the key thematic layers need to be identified
	3. The scale ranges and spatial representations for each thematic layer need to be specified.
	4. The data that would be represented should be grouped into datasets for effective orderly representation of the data.
Logical design	5. The tabular database structure and behaviour for descriptive attributes need to be identified.
	6. Spatial properties of your datasets need to be defined.
	7. Propose a geodatabase design.
Physical design	8. Implement, prototype, review and refine the design.
	9. Design workflows for building and maintaining each layer.
	10. Document the design using appropriate methods.

Table 3.2. Ten steps to designing a geodatabase by Arctur & Zeiler (2004)

3.2. Applying the geodatabase design

A clear and thorough understanding of the capabilities and structure in which GIS operates and stores information need to be understood in order to understand the procedures that should be followed for integrating the water network system with GIS. Emphasis on

understanding the structure of the water network should also be stressed. Having a holistic understanding of the two environments would provide a more accurate approach to the geodatabase design and the way that GIS could be applied for storing, managing and representing information. With this in mind and applied, the ten steps towards creating a geodatabase as indicated in Table 3.2 are discussed.

3.2.1. The information products that will be produced with the GIS

This step contains the information that had to be created for the project. It was therefore important to start with the end goal in mind in order to identify what information would serve as the building blocks for the structure of the geodatabase. Firstly, and fore mostly, it was important to keep in mind that the focus of data used in the geodatabase is spatially related.

With reference to Chapter 1, the aim of the study was to establish a prototype geodatabase that would serve as a container for the water network data of buildings E4 and E6. The geodatabase would therefore also serve as an inventory for selected water utilities of the proposed section on the campus. In other words, inventory map products had to be created with the relevant information regarding potable water supply for the sections of buildings E4 and E6. In order to determine what information would be used in the database, it was important to study the current paper and CAD drawing plans. A generic set of features had to be created based on the current data available. Furthermore, with the use of GIS, the possibility of representing the data in two- and three-dimensional views provided an indication of the data that had to be collected. Editing and updating of information have to be easy and therefore digitizing the CAD data from original CAD drawings was done in order to provide spatial referenced drawings and representations of the relevant data.

Firstly the georeferenced CAD files have to be created. Subsequently the features within the CAD files have to be digitized in order to create different feature classes. A feature dataset with the correct reference system have to be established. This would ensure that the feature classes are stored in an organized "directory" as organized entities. In order to provide additional descriptive information for the feature classes, attribute fields have to be created. Supporting the data integrity of the digitized data and entities of the feature classes, topology rules, subtypes, domains and relationships between the different feature classes have to be established.

Following the creation of a suitable environment for the data, the development for a proposed 3D model of buildings E4 and E6 could be established. Based on the existing CAD data, a complete potable water network system have to be created. The network should consist of points, lines and polygons, each contributing to the holistic extent of the proposed water network model for buildings E4 and E6. The model has to serve as an information management system for the water network within the proposed buildings with information of each and every entity along the water lines. This way the retrieval of information regarding the components would be simplified. The geodatabase would serve as a central container for all the relevant data concerning the potable water network system of buildings E4 end E6. Additional information with regard to the maintenance records have to be contained within the geodatabase. The final product of the "PUK_Kampus5" geodatabase should serve as a management system for technical staff. Information within the geodatabase would be available to simplify the work executed on relevant water lines. For the purpose of this study and due to technical reasons, there will be referred to the geodatabase created for this study as the "PUK_Kampus5" geodatabase.

3.2.2. Identifying the key thematic layers

Following step one the information requirements that are necessary for the data model had to be as accurate as possible. At this stage it was clear that not all the data sources with regard to the study area were up to date. This implied that accuracy of the data were lacking. However, relevant data concerning the study area were collected and included a QuickBird image (2008) with a Universal Transverse Mercator projection and a spatial reference of WGS 1984 UTM Zone 35S. This simplified the projection transformation for the datasets within the geodatabase as well as all the additional information that would be added to the geodatabase. The CAD drawings used in the design of the geodatabase are presented on a scale of 1:100. Some attribute tables have been created inside the geodatabase as standalone tables and contains information with regard to the owners of each room. The same attribute tables were created for the contractors that are listed as the worker force for each utility. An attribute list of all the facilities in the system and a list of all the maintenance information along the waterlines were also created. In order to represent a complete view of the data inside the database, a composition of the basic map layers, chosen to represent the project outline, were added.

These layers are briefly mentioned and more carefully discussed in step 6 as presented in Table 3.2:

3.2.2(a) *Layer:* The management zones in which the campus is divided;

3.2.2(b) *Layer:* A building outline of building E4 and E6;

3.2.2(c) *Layer:* PUK rooms;

3.2.2(d) *Layer:* End point fittings;

3.2.2(e) *Layer:* The fittings for all the waterlines;

3.2.2(f) *Layer:* Geysers;

3.2.2(g) *Layer:* Mains;

3.2.2(h) *Layer:* Manhole (Chambers);

3.2.2(i) *Layer:* Pumps;

3.2.2(j) *Layer:* Meters;

3.2.2(k) *Layer:* System valves;

3.2.2(l) *Layer:* Control valves;

3.2.2(m) *Layer:* Thrust protection;

3.2.2(n) *Layer:* QuickBird image (2008); and

3.2.2(o) *Layer:* Fishnet.

The logical design

3.3. The scale ranges and spatial representations for each thematic layer

Table 3.3 represents a summarized table of the layers that are proposed for the geodatabase together with information regarding the description for each layer as well as the data type, scale and spatial representation. The data used for the study are presented on a scale of 1:100.

Layer name	Description	Data type	Scale	Spatial representation
PUK_Zones	A polygon representation of the management zones of the campus	Vector	1:100	Polygons
PUK_Buildings	A polygon representation of the building outlines of all the buildings	Vector	1:100	Polygons
PUK_Rooms	A polygon representation of all the rooms in the buildings	Vector	1:100	Polygons
End point fittings	Points representing all the fittings at the end nodes of all the edges	Vector	1:100	Points
Fittings	Points representing all the fittings in the network system	Vector	1:100	Points
Pumps	Point representations of the pumps in the water network system	Vector	1:100	Points
Geysers	Point representations of the geysers in the system	Vector	1:100	Points
Kitchen boilers	Point representations of the kitchen boilers in the system	Vector	1:100	Points
Meters	Point representations of all the meters in the system	Vector	1:100	Points
Manholes (Chambers)	Point representations of all the manholes in the area	Vector	1:100	Points
Mains	Line representations of all the water provision lines inside the buildings	Vector	1:100	Lines
Trust protection	Points representing the corner protection for the water lines	Vector	1:100	Lines

System valves	Point representations of the systems valves in the network	Vector	1:100	Points
Control valves	Point representations of the system valves in the network	Vector	1:100	Points
QuickBird image (2008)	A raster image representing an aerial picture of the study area	Raster	1:100	Pixels
Fishnet	A gridded layer covering the study area in order to serve as a medium of orientation for facilities outside buildings	Vector	1:100	Gridded lines

Table 3.3. The scale ranges and the spatial representation for each layer

3.4. Group representations into datasets

3.4.1. Feature datasets

The feature classes were organized in feature datasets that consist of a spatial reference system GCS WGS 1984 and a datum: WGS 1984. All the feature classes within the feature dataset adopted the specified reference system which contributed greatly to the accuracy of the data.

3.4.2. Feature classes

The feature classes are those files that are represented as layers, each containing and representing a different feature of the water network. The following feature classes, represented in Tables 3.4 up to Table 3.16 are located and maintained inside the "Water" feature dataset within the PUK_Kampus geodatabase.

A single feature dataset contains all the different feature classes in range and also contains the topology feature class and the network dataset. The topology feature class and the network dataset serve as supportive mediums containing rules and behaviour to regulate and validate the legitimate creation of features inside the geodatabase. A more descriptive explanation is that the feature classes within the feature dataset have consistent rules established for modelling the network and joining the features into a continuous network operating system.

Relationships were also established between the features which are represented as feature classes among the standard features in the dataset.

Some features are not located within the feature dataset and serve as stand-alone features located inside the geodatabase. This includes tables providing information with regard to the network such as the "Owners table", the "Maintenance record table" and the "List of contractors table". The reason for this occurrence is that these tables do not have any spatial properties to enable their location in the system. Furthermore, a raster dataset such as the QuickBird image (2008) of the study area, are also located as a loose standing data source because it serves mainly as an information given layer not taking part in the network.

3.5. Tabular database structure for attributes

This step covers a more detailed exemplification of the attribute fields and valid values and ranges that apply to these attribute fields. This includes a description of the tables, subtypes, domains, topology rules and so forth.

3.5.1. Feature classes

The delimitation of the campus in different management zones are represented in Table 3.4. The zones range from zones A – L with the focus primarily on zone E as the study area for buildings E4 and E6. The zones will be represented as polygons and will contain the infrastructure. These zones were digitized from CAD drawings.

PUK_Zones (Polygons)			
Field name	**Data type**	**Allow nulls**	**Domain**
OBJECTID	**Object_ID**		
SHAPE	**Geometry**	**Yes**	
Unique_ID	**Text**	**Yes**	
Number_of_Buildings	**Long Integer**	**Yes**	
SHAPE_Length	**Double**	**Yes**	

Table 3.4. PUK_Zones feature class

The building outlines (Table 3.5) serve as the structure in which the rooms inside the buildings are located. These building outlines were also digitized from original CAD data which were integrated into GIS.

PUK_Buildings (Polygons)			
Field name	**Data type**	**Allow nulls**	**Domain**
OBJECTID	Object_ID		
SHAPE	Geometry	Yes	
Unique_ID	Text	Yes	
Zone_ID	Long Integer	Yes	
Name_of_Building	Text	Yes	
Faculty	Long Integer	Yes	Faculty
Department	Long Integer	Yes	Department
Number_of_Floors	Long Integer	Yes	
SHAPE_Length	Double	Yes	
SHAPE_Area	Double	Yes	

Table 3.5. PUK_Buildings feature class

Manhole chambers (Table 3.6) are sections along the water pipeline with access points towards features such as valves, meters or pumps for inspection purposes. The size of these access points are much smaller than those of a manhole but serves the same purpose on a smaller scale. Each chamber is represented by a point in the network system.

Manhole_Chambers (Points)			
Field name	**Data type**	**Allow nulls**	**Domain**
OBJECTID	Object_ID		
SHAPE	Geometry	Yes	
Unique_ID	Text	Yes	
Additional_Comments	Text	Yes	
Main_Section_Connected	Text	Yes	

Table 3.6. Manhole_Chambers feature class

The PUK_Rooms feature class (Table 3.7) represents the rooms inside each building digitized from the original CAD data on four levels including the underground (sub-level). The rooms on all the levels have been indicated; sub-level, level one, level two and level three together with the stairs inside the buildings. The rooms are represented by polygons and

in general these rooms are the spaces containing most of the facilities in the water network system.

PUK_Rooms (Polygons)				
Field name	**Data type**	**Allow nulls**	**Domain**	**Subtype**
OBJECTID	Object_ID			
SHAPE	Geometry	Yes		
Unique_ID	Text	Yes		
Building_ID	Long Integer	Yes		
Space_Type	Long Integer	Yes	Space_Type	
Owner_ID	Long Integer	Yes		
Capacity	Long Integer	Yes		
Floor	Long Integer	Yes		Floor
SHAPE_Length	Double	Yes		
SHAPE_Area	Double	Yes		

Table 3.7. PUK_Rooms feature class

A feature class indicated by Table 3.8 contains information about the thrust protection in the network, represented by points. The thrust protection is concrete blocks placed at strategic turns in the water network and serve as a medium to protect the network equipment from moving or rupturing during sudden changes in pressure.

Thrust_Protection (Points)				
Field Name	**Data type**	**Allow nulls**	**Domain**	**Subtypes**
OBJECTID	Object_ID			
SHAPE	Geometry	Yes		
Unique_ID	Text	Yes		
Install_Date	Date	Yes		
Installed_by_Contractor	Long Integer	Yes		
Trust_Protection_Type	Long Integer	Yes		Trust_Protection_Type

Table 3.8. Thrust protection feature class

The mains layer (Table 3.9) represents the two major water distribution mains: *Distribution mains and the Lateral mains.* In order to establish the correct topological rules for each of

these line features, these two features have been grouped together. The feature class contains relevant data and attribute data for each of these network lines.

The lateral mains provide water towards the different rooms: hot and cold. This layer contains all the different waterlines connected from the main distribution mains towards the endpoints in the water network system. These lines would also be represented on different elevation heights with reference to the buildings.

Major distribution mains distribute water from municipal sources such as reservoirs and feeds water towards the lateral mains. Such as all the different layers, the source of the main water distribution lines is CAD data.

Mains (Lines)				
Field name	**Data type**	**Allow nulls**	**Domain**	**Subtypes**
OBJECTID	**Object _ID**			
SHAPE	**Geomtry**	**Yes**		
Unique_ID	**Text**	**Yes**		
Diameter	**Long Integer**	**Yes**	**Fitting_Diameter_Size _mm**	
Material	**Text**	**Yes**	**Material**	
Maximum_Pressure_Rating	**Text**	**Yes**		
Water_Temp	**Long Integer**	**Yes**	**Water_Temp**	
Main_Subtype	**Long Integer**	**Yes**		**Main_Subtype**
SHAPE_Length	**Double**	**Yes**		
Installed_by_Contractor	**Long Integer**	**Yes**		

Table 3.9. Water mains feature class

The geysers feature class (Table 3.10) contains information about all the geysers in the network as well as descriptive data and subtypes which are all represented by point features. The geysers and the kitchen boilers were merged into one single feature class. The reason for the combination of the two features is subjected to the creation of domains that would eliminate the need to have two separate feature classes. The kitchen boilers are classified as a geyser type and therefore had a special domain created for it. The combination of the two features brought about the elimination of another feature class which contributed to the limitation of unnecessary feature classes.

Geysers (Points)			
Field name	**Data type**	**Allow nulls**	**Domain**
OBJECTID	Object_ID		
SHAPE	Geometry	Yes	
Room_ID	Long Integer	Yes	
Unique_ID	Text	Yes	
Install_Date	Date	Yes	
Installed_by_Contractor	Long Integer	Yes	
Status	Long Integer	Yes	Status
Geyser_Capacity	Long Integer	Yes	Geyser_Size_Litres
Geyser_Element_Rating	Long Integer	Yes	Geyser_Element_Rating
Geyser_Operating_Pressure__kPa	Long Integer	Yes	Operating (kPa)
Geyser_Mass_kg	Text	Yes	Geyser_Mass_(kg)
Geyser_Code	Text	Yes	Geyser_Type_Code
Main_Sections_Connected	Text	Yes	

Table 3.10. Geysers feature class

The pumps layer (Table 3.11) represents all the pumps in the system that are applied for enhancing the pressure in the system whenever the need for increased water pressure occurs or to introduce additional flow into the network.

Pumps (Points)				
Field name	**Data type**	**Allow nulls**	**Domain**	**Subtypes**
OBJECTID	Object_ID			
SHAPE	Geometry	Yes		
Unique_ID	Text	Yes		
Inlet_Diameter	Long Integer	Yes	Fitting_Diameter_Size _mm	
Discharge_Diameter	Long Integer	Yes	Fitting_Diameter_Size _mm	
Install_date	Date	Yes		
Installed_by_Contractor	Long Integer	Yes		
Status	Long Integer	Yes	Status	
Pump_types	Long Integer	Yes		Pump_types
Main_Sections_Connected	Text	Yes		

Table 3.11. Pumps feature class

Point features (Table 3.12), represents control valves which are the valves that could be located anywhere in the system in order to control the water flow with special operating characteristics such as backflow control, air control, air altitude etc.

Control valve (Points)				
Field Name	**Data Type**	**Allow Nulls**		**Subtypes**
OBJECTID	Object_ID			
SHAPE	Geometry	Yes		
Diameter	Long Integer	Yes	Fitting_Diameter _Size_mm	
Water_Temp	Long Integer	Yes	Water_Temp	
Unique_ID	Text	Yes		
Install_Date	Date	Yes		
Status	Long Integer	Yes	Status	
Control_Valve_Type	Long Integer	Yes		Control_Valve_Type
Installed_by_Contractor	Long Integer	Yes		
Material	Text	Yes	Material	
Main_SectionsConnected	Text	Yes		

Table 3.12. Control valve feature class

Water meters (Table 3.13) serve as devices that are used to measure the volume of water usage.

Meters (Points)			
Field Name	**Data Type**	**Allow Nulls**	**Domain**
OBJECTID	Object_ID		
SHAPE	Geometry	Yes	
Room_ID	Long Integer	Yes	
Unique_ID	Text	Yes	
Diameter	Long Integer	Yes	Fitting_Diameter_Size_mm
Water_Temp	Long Integer	Yes	Water_Temp
Install_date	Date	Yes	
Installed_by_contractor	Long Integer	Yes	
Status	Long Integer	Yes	Status
Main_Section_Connected	Text	Yes	

Table 3.13. Meters feature class

These meters are used in residential and commercial buildings in public water supply system. Water meters could be used at the water source, well or throughout the water system in order to determine the flow rate through that portion of the system. Water measurement and total usage are displayed in cubic feet (ft³), cubic meters (m³) or US gallons on a mechanical or electronic register. These meters are used to measure the flow of the water through the network and are also used for billing purposes.

A feature class (Table 3.14) represents the endpoint fittings for the following features: Toilet water inlets, urinal water inlets and taps (including taps for showers and bath tubs). All of these features occur at the endpoints of the water lines and were therefore placed in a single feature class. Different end point fittings serve as subtypes in this feature class.

End point facilities (Points)				
Field name	**Data type**	**Allow nulls**	**Domain**	**Subtypes**
OBJECTID	**Object_ID**			
SHAPE	**Geometry**	**Yes**		
Unique_ID	**Text**	**Yes**		
Room_ID	**Long Integer**	**Yes**		
Install_Date	**Date**	**Yes**		
Installed_by_Contractor	**Long Integer**	**Yes**		
Status	**Long Integer**	**Yes**	**Status**	
Water_Temp	**Long Integer**	**Yes**	**Water_Temp**	
Facility_Subtype	**Long Integer**	**Yes**		**Facility_Subtype**
Material	**Text**	**Yes**	**Material**	
End_Point_Diameter	**Long Integer**	**Yes**	**Fitting_Diameter_Size_mm**	
Main_Section	**Long Integer**	**Yes**		

Table 3.14. End point facilities feature class

The system valve layer (Table 3.15) contains point features, representing system valves that regulates the pressure, isolates portions of the network, throttles flow, prevents backflow and relieves pressure.

System_Valve (Points)				
Field name	**Data type**	**Allow nulls**	**Domain**	**Subtypes**
OBJECTID	Object_ID			
SHAPE	Geometry	Yes		
Motorized	Long Integer	Yes	Yes/No	
Percent_open	Long Integer	Yes	Percent_Open	
Pressure_Setting	Text	Yes		
Diameter	Long Integer	Yes	Fitting_Diameter_Size _mm	
Unique_ID	Text	Yes		
Room_ID	Long Integer	Yes		
Water_Temp	Long Integer	Yes	Water_Temp	
Install_Date	Date	Yes		
Installed_by_Contractor	Long Integer	Yes		
Status	Long Integer	Yes	Status	
System__Valve_Subtypes	Long Integer	Yes		System__Valve _Subtypes
Main_Sections_Connected	Text	Yes		

Table 3.15. System valve feature class

The fittings layer (Table 3.16) represents all the point/end nodes along the edges of the water network lines that contain fittings such as bends, caps, couplings, risers, reducers, sleeves, saddles etc. These fittings are used to link and join different sections of the water network pipes.

The QuickBird layer is a digital aerial photo which is already spatially referenced and used as the backdrop photo as reference for the CAD data that are imported into ArcGIS. The QuickBird photo (2008) serves as the main source of visual reference for the system.

In order to have a medium of reference for the facilities outside the buildings a layer called the "fishnet" was created. It is gridded layer consisting out of grids (blocks) with a 1 meter × 1 meter size. The location of a facility outside a building could be located with reference to the fishnet blocks. The grid sizes could also be adjusted and resized in order to provide a more or less accurate way of orientation.

Fittings (Points)				
Field name	**Data type**	**Allow nulls**	**Domain**	**Subtypes**
OBJECTID	**Object_ID**			
SHAPE	**Geometry**	**Yes**		
Unique_ID	**Text**	**Yes**		
Install_Date	**Date**	**Yes**		
Installed_by_Contractor	**Long Integer**	**Yes**		
Material	**Text**	**Yes**	**Material**	
Type	**Long Integer**	**Yes**		**Type (Fittings)**
Fitting_Diameter	**Long Integer**	**Yes**	**Fitting_Diameter_ Size_mm**	
Angle_Degrees	**Long Integer**	**Yes**	**Angles_Degrees_**	
Water_Temp	**Long Integer**	**Yes**	**Water_Temp**	
Main_Sections_Connected	**Text**	**Yes**		

Table 3.16. Fittings feature class

3.5.2. Tables

Tables that provide static and adjustable information for the geodatabase are represented as stand-alone tables inside the geodatabase. This means that the tables are not part of the feature dataset and are located outside of the feature dataset. The tables contain no spatial represented data and are merely information based. These tables are:

- *List of contractors:* Table 3.17 represents a static table that contains information about the contractors generally approached for any type of maintenance purposes of the water network system. The table includes the following information about each contractor: name, occupation (whether it is a plumber, electrician, or IT technician), contact details, the company and the campus that the contractor is active at.

List_of_Contractors		
Field name	**Data type**	**Allow nulls**
OBJECTID	**Object_ID**	
Contractor_Name	**Text**	**Yes**
Occupation	**Text**	**Yes**
Contractor_Tel_nr	**Text**	**Yes**
E-mail	**Text**	**Yes**
Company	**Text**	**Yes**
Campus	**Text**	**Yes**
Fax_nr	**Text**	**Yes**

Table 3.17. List of contractors table

- *Owners:* Table 3.18 represents a static table that contains information about any person that might be in charge of a specified area or room in the buildings as well as the persons who occupy offices inside the buildings. This table is used as a reference list of people that could be contacted whenever a particular maintenance task should be executed in a specific room or area within a building. This table contains the following fields: Object_ID, Owner_ID, Owner_Name, Tel_Number, E-mail and Department.

Owner_Table		
Field Name	**Data Type**	**Allow Nulls**
OBJECTID	**Object_ID**	
Owner_ID	**Long Integer**	**Yes**
Owner_Name	**Text**	**Yes**
Tel_Number	**Text**	**Yes**
E_mail	**Text**	**Yes**
Department	**Long Integer**	**Yes**

Table 3.18. Owner table

- *Maintenance record:* Table 3.19 represents an adjustable table that contains information frequently adjusted and updated. The purpose of this table is to store and keep record of updated information about maintenance procedures executed over time. The importance of this table could not be overstated and includes the following

fields: Object Identifier, Job_ID, Job_Type, Job_Description & Comments, Contractor_ID, Date_of_Service, Due_Date, Location_Description, the Duration_of_Maintenance and Facility_Serviced. Updating information prevents confusion and uncertainty about the water network system and its functions.

Maintenance_Record		
Field Name	**Data Type**	**Allow Nulls**
OBJECTID	**Object_ID**	
Job_ID	**Long Integer**	**Yes**
Job_Type	**Long Integer**	**Yes**
Job_Description_Comments	**Text**	**Yes**
Contractor_ID	**Long Integer**	**Yes**
Date_of_Service	**Date**	**Yes**
Due_Date	**Date**	**Yes**
Location_Description	**Text**	**Yes**
Maintenance_Duration	**Text**	**Yes**
Facility_Serviced	**Text**	**Yes**

Table 3.19. Maintenance record

3.5.3. The Unique Identifier

A special identifier was created for the unique identification of each feature. Even though each feature has a special unique "Obect_ID created for it, the "Unique_ID" created for each feature serves as a cryptic code that contains information about each aspect of that specific feature. Each "Unique_ID" are divided in a standard set of specified units represented in Table 3.20. With reference to the "Unique_ID" for a certain feature, a considerable amount of information with regard to the feature could be derived. It also simplifies the process of selecting a query or searching for a certain feature by means of "Select by Attributes" in the main toolbar. An example of the composition for the "Unique_ID" ZE/E6/FS1/bend/70 is indicated in Table 3.20.

	Description	Example
Zone	The zones could be anything from Zone A up to Zone L – all the management zones that the campus is divided into	ZE = Zone E
Building Code	The building code is the codes used by the university to indicate the zone in which the building is located as well as the number of the building on the campus	E6 = Building 6 located in Zone E
Floor Level	Whenever a feature are located within a certain floor its "Unique_ID" would represent the floor level	FS1 = Floor Sub-level 1
Feature Type	The feature type would be specified	bend or cross or tee
Object_ID	The original "Object_ID" for the feature would be represented	Object_ID = Object identifier 70

Table 3.20. A complete description of the "Unique_ID"

3.5.4. Subtypes

Up till now, it was found that subtypes serve as a storage mechanism for features, their entities and attribute data and could be specified for each feature class. Subtypes are predefined types for a certain feature. Unlike domains, subtypes also have the ability to contain more descriptive data for the feature it is created for.

During the study notice were taken of the fact that different features should not be added as subtypes in one feature class. Whenever the features' fields correspond with each other, subtypes could be created. However, whenever the fields between different entities differ too much, it would rather be more appropriate to create different feature classes for each feature separately. Subtypes of each feature could be created as these are actually subtypes from the feature itself and do share the same fields with little differences between the entities. Deciding whether a subtype or domain should be added depends on the application possibilities of the subtype or domain. A subtype is only applicable to the feature for which it is specified for and not destined to serve as a predefined entity that could be used right across the geodatabase for any feature such as a domain. A constant and general concept that summarizes the aforementioned; a subtype could be created from a feature but an

independent feature could not serve as a subtype except if it shares the same fields. It should be stressed that the data type of a subtype is a long integer and should correspond with the same data type in the specified attribute field. Referring to Addendum C, the subtypes for certain features are represented in the blue tables. A listed view of the subtypes is available in Addendum A.

3.5.5. Domains

Data integrity and accuracy is one of the aspects that cannot be overemphasized. Using predefined attribute values for different features ensures that the data entered are valid, accurate and correctly spelled. Features might have a few different valid values as in the case of a pipeline. The pipeline could have diameters of different sizes at certain sections or it could exist out of different material types. The options for these entities do not need to be explained by means of extra additional attribute data such as subtypes or feature classes but could serve as a predefined option that forms part of the bigger defined attribute data for a certain feature. The domains are created inside the geodatabase and could be applied in any feature classes located inside the geodatabase with the same data type. Once a domain has been created it cannot be removed from the geodatabase. Domains should therefore be created based on good planning and should also selectively be added to the geodatabase. Referring to Addendum C, the domains created for the geodatabase are indicated in red. A listed view is available in Addendum A and for a more distinct version of Addendum C please refer to the electronic format provided thereof.

3.6. Define the spatial properties of the datasets

3.6.1. Topology rules

Topology rules were created in order to enforce integrity between the features and how they interact with each other. Utilizing topology in the network between the features would restrict the possibility of creating features with errors but rather promote legal connectivity creation of the features. The following topology rules used in this system are applied to the features:

- end point facilities *Must be covered by endpoint of* Mains;
- fittings *Must be covered by endpoint of* Mains;
- geysers *Must be covered by endpoint of* Mains;
- mains *Must be inside* Zones;

- meters *Must be covered by endpoint of* Mains;

- pumps *Must be covered by endpoint of* Mains;

- system valve *Must be covered by endpoint of* Mains;

- thrust protection *Must be covered by endpoint of* Mains;

- PUK buildings *Must not overlap;*

- PUK zones *Must not overlap; and*

- control valve *Must be covered by endpoint of* Mains.

Topology rules represented:

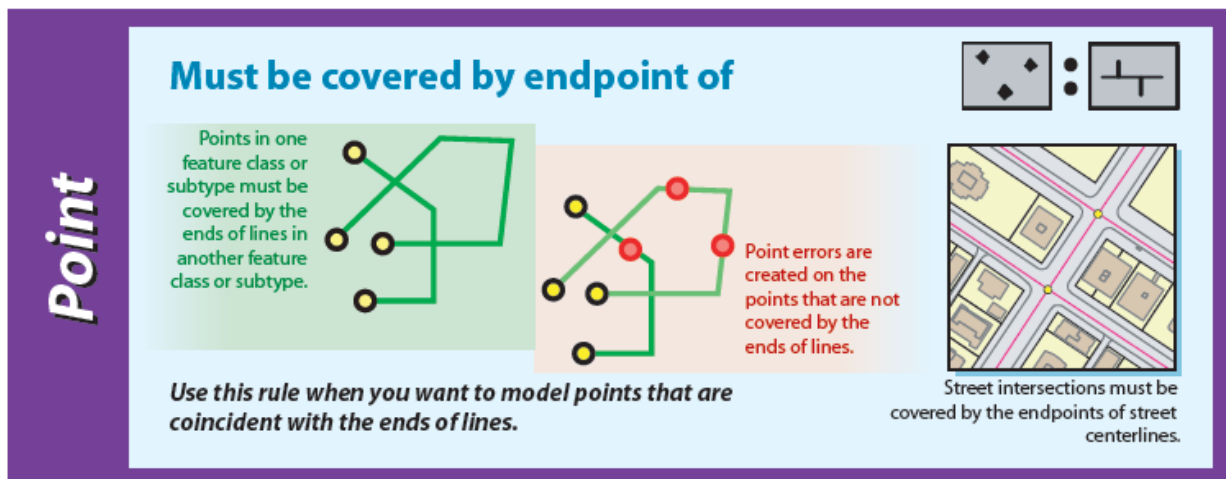The three topology rules used in the geodatabase are illustrated with Figures 3.2, 3.3 and 3.4 below:

Figure 3.2. Topology rule defined between points and lines
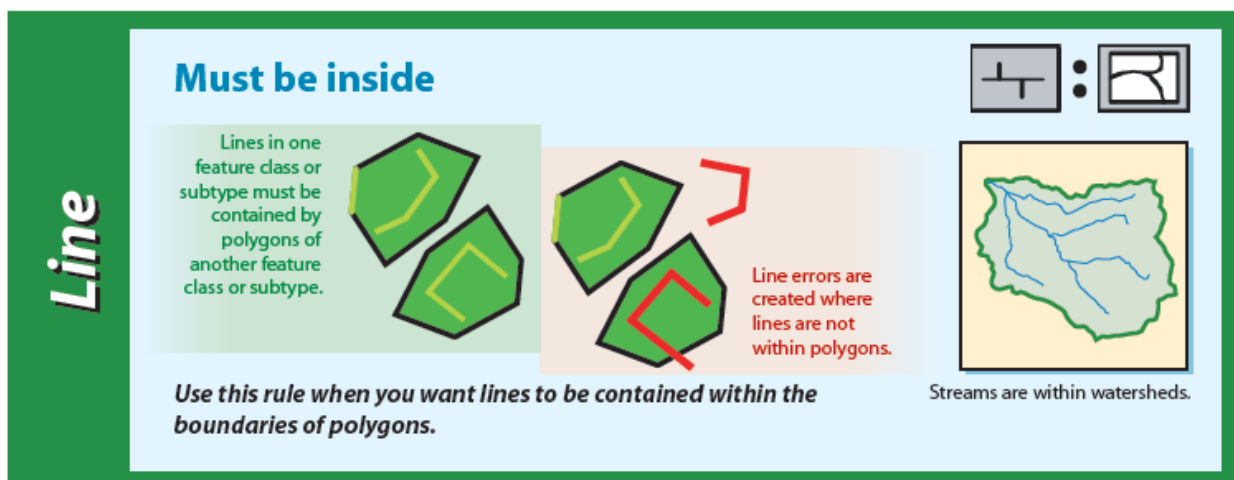
Figure 3.3. Topology rules established for the polygons in the system
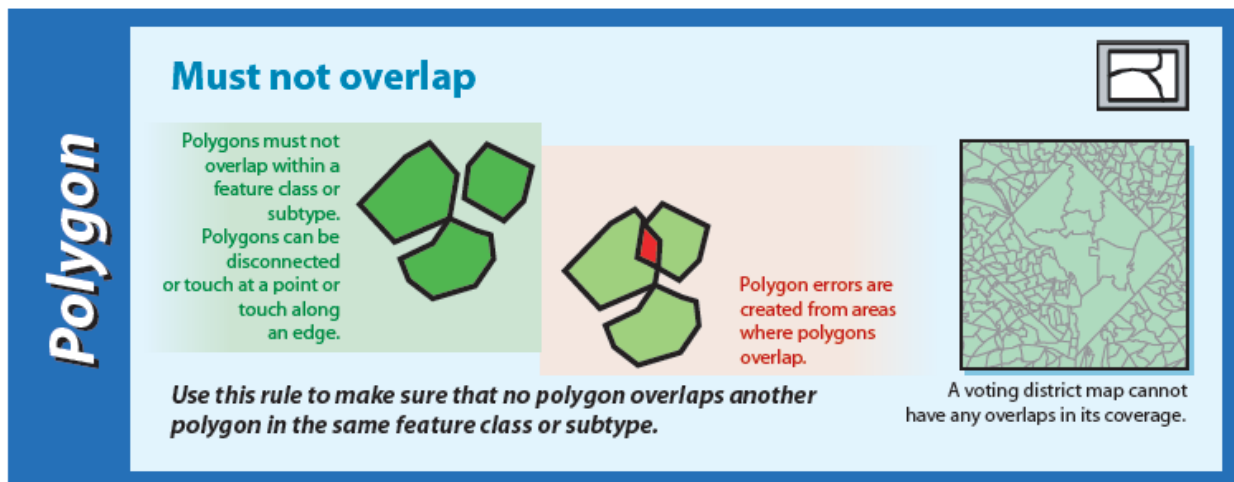
Figure 3.4. Topology rule established between lines and polygons

Referring to the three topology rules indicated in Figures 3.2, 3.3 and 3.4 the points, lines and the polygons have topology rules defined between them. All the points must be covered by the endpoint of the mains, all the polygons must not overlap and the mains (lines) *"must be inside"* the PUK zones polygons. The reason for this is that the distribution mains as well as the lateral mains are located inside and outside the buildings and the best topology rule defined for them are the one that supports this concept and provides the establishment of new features within and outside the PUK Buildings. Lateral mains and distribution mains are both located inside the one feature class called "Mains", which also motivates this selection of topology.

All the point features that participate in the topology rules have the rule *"must be covered by endpoint of"* Mains, specified for them. The reason for applying this rule is that all the lines are created in sections between points. This means that each and every section of the lines is created between point features which imply that there are no continuous lines crossing any point at any time. This concept is therefore supported by the topology rule defined for all the end points. For each and every polygon feature, the rule *"must not overlap"* has been defined to ensure that none of the polygons overlap each other.

In order to establish correct and accurate topology between the features the topology building process were repeated until the correct error free topology rules were selected for each feature participating in the topology dataset. Some of the features with topology errors were adjusted to obey the topology rule created for it. A more detailed theoretical description of topology is indicated in section 2.5.1.4, Chapter 2.

3.6.2. Applying spatial reference

With reference to section 4, Chapter 2, the different spatial references that were considered have been summarized in Table 2.11. It was found that the Universal Transverse Mercator, WGS 1984 UTM Zone 35S is the spatial reference system selected due to its accuracy, versatility and minimal distortion of scale within the entire zone. Furthermore, the QuickBird image (2008) is also projected accordingly which simplifies the transformation process for the rest of the features. Whenever a co-ordinate system had to be defined for a feature dataset or feature classes, the defined co-ordinate system have been imported from the QuickBird image (2008). Working with one predefined accurate co-ordinate system ensures accuracy throughout the entire geodatabase.

3.7. Propose a geodatabase design

3.7.1. Relationship classes

Relationships between the different feature classes and the aforementioned tables were established in order to establish data integrity and to exclude repetition of fields inside the geodatabase. The data types of the fields that are related are one of the concepts that were emphasized. E.g. long integer fields could only be related with a field that also has a long integer as a data type. This same principle applies for text, date/time and the rest of the data types.

Indicated in Figure 3.5 are the feature classes represented in orange (the physical infrastructure). The blue feature classes represent the stand-alone tables containing information with regard to the system and the green feature classes represent the components of the water network.

Following a series of attempts executed for creating the correct relationship classes, the final layout of the relationship classes between the features has been established with an indication of the cardinality between the relationship classes presented. The procedure entailed a time intensive process of creating the features in ArcMap after which they were then analyzed and related to each other according to the corresponding fields in each feature class. The features connecting directly on the water mains all have the foreign key "Main_Section_Connected" which are the corresponding field that enables the creation of relationships between the features.
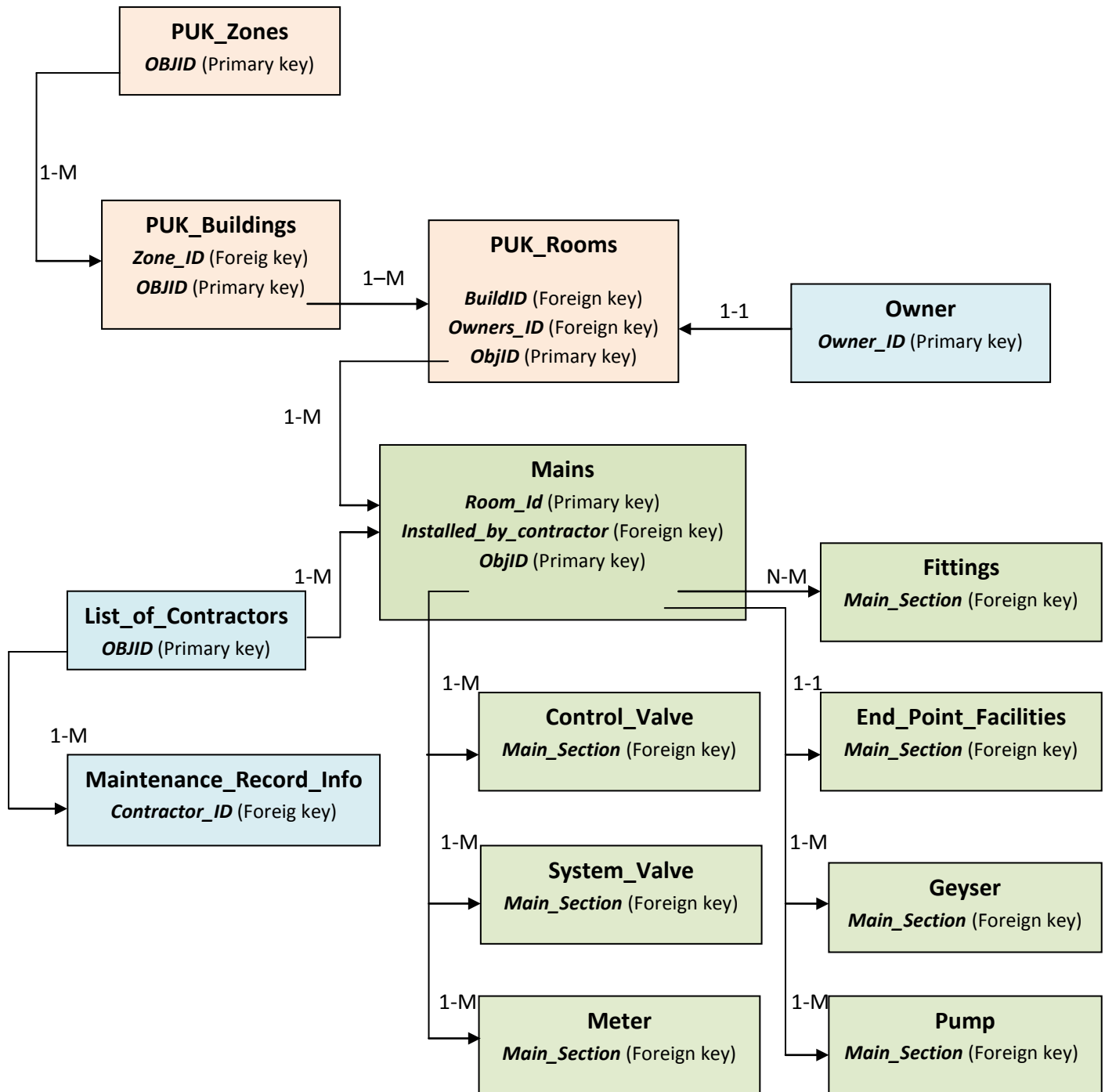
Figure 3.5. The relationship classes together with their cardinalities

A one-to-one relationship has been established between the features that share literally one feature. This is especially the case with the end point facilities due to the fact that each section of the water mains between the points represents a separate feature with a unique OBJECTID as represented in Figure 3.6. The one-to-one relationship cardinality is represented with the connection between line L7 and point P5. Each of these only has one

feature type connected to one another and therefore a one-to-one relationship is established. An example is the end point fittings at the end of each line.
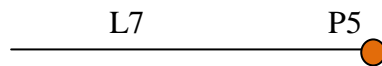

Figure 3.6. The one-to-one relationship cardinality

A many-to-many relationship could be established as indicated in Figure 3.7. This implies that whenever two of the same features are located at each end of the line, the many-to-many relationship is established. Points P2 and P3 both share the line section L2 and each point also shares in the line section L1 and L3 separately.
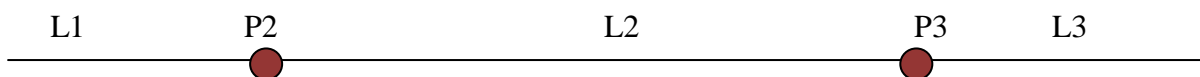

Figure 3.7. The many-to-many relationship representation

The following many-to-many relationship could therefore be established (Table 3.21):

OBJID	Line	Point
1	L1	P2
2	L2	P2
3	L2	P3
4	L3	P3

Table 3.21. A demonstration of the many-to-many relationship classes

Whenever two features are shared by one mutual feature a one-to-many cardinality are represented (Figure 3.8):
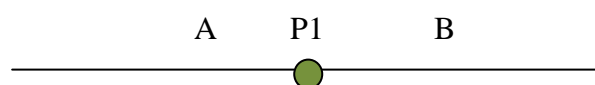

Figure 3.8. A one-to-many relationship

Line sections A and B in Figure 3.8 are individual sections of the network line, and meets up at point P1. This indicates that point P1 shares both lines A and B which implies a one-to-many relationship.

The thrust protection and manhole feature classes were not added to the relationship model because of their fixed and invariable operating nature in the network system.

The feature classes, tables, domains and subtypes presented in this Chapter is the result of substantial prototypes that were tested and refined before they were finalized. Within Chapter 3 a brief discussion on the conceptual and logical design steps of the geodatabase were provided. The components that form part of the design were also examined and discussed more closely. The initial purpose of Chapter 3 was to indicate how the conceptual and logical design stages of the geodatabase are created and how each step contributes to the general design of the proposed geodatabase. Within Chapter 3 the transition between the theoretical background and the actual design of the geodatabase are indicated. The logical and conceptual design phases therefore serve as the platform for the physical design stage that would be based on the proposed ideas and concepts for the features that would physically be applied in Chapter 4. Additional information of the system and its components were also provided in order to prevent any misconceptions as far as possible.