

**Creating a seamless geodatabase for water infrastructure on the  
Potchefstroom campus.**

A. L. du Toit  
20076185

Dissertation submitted in partial fulfilment of the requirements for the degree  
*Master of Science*  
at the Potchefstroom Campus of the North West University

Supervisor: T. C. De Klerk

April 2012

## **ACKNOWLEDGEMENT**

Firstly I would like to thank my Heavenly Father who enabled me to finish this thesis to the best of my ability and provided me with the skills and privilege to execute this study.

I am also heartily thankful to my supervisor, Theuns de Klerk, whose encouragement, guidance and support from the initial to the final level enabled me to develop an understanding of the subject and the final product of the study.

I would like to thank my colleagues, Carl Bester and Dawie Maree, who provided their support and great co-operation during the course of this study.

Lastly I offer my regards and blessings to all of those who supported me in any regard during the completion of the project.

Armand du Toit

## **ABSTRACT**

The Potchefstroom Campus of the North West University contains old water pipelines that are not well documented. Many of the newer water pipelines are not well documented either. A central data storage system that could contain the information with ease of access to update and retrieve information of these waterlines is lacking. There is a need to find a way that existing potable water network data could be represented and stored with GIS. The solution would contribute to the management of the water system on Campus.

The aim of this study is to create a seamless geodatabase as a pilot project for the potable water infrastructure at the Potchefstroom Campus of the North West University. The pilot project focuses on buildings E4 and E6. ArcGIS 10 was selected to serve as the key software system that would be applied as a medium to solve and represent the problem. ArcGIS geodatabase serves as a container to store spatial data with. Data with regard to the potable water system was collected from various sources of which available electronic and hard copy CAD data was the general format.

A file geodatabase was created in ArcCatalog with a standard co-ordinate system as reference to the data. ArcMap was applied for 2D editing and georeferencing of the CAD drawings which were followed by a composition of attribute data for the created features. The end result was represented in ArcScene for 3D visualization and 3D analysis. It also provided ease of access to the attribute information and relationships and the capability to perform the shortest route analysis.

Keywords: GIS, geodatabase, water infrastructure, 3D, network

## OPSOMMING

Die Potchefstroom Kampus van die Noord-Wes Universiteit bevat ou water pype wat nie goed gedokumenteer is nie. Baie van die nuwer pyplyne word ook nie goed gedokumenteer nie. 'n Sentrale sisteem wat kan dien as 'n stoorplek vir die data ontbreek tans. Hierdie data sisteem sal ook gemaklike toegang tot die data moet bied van waar dit opgedateer kan word. Daar is tans 'n behoefte om 'n manier te vind waarmee die bestaande water netwerk data voorgestel en gestoor kan word. Die oplossing sal dus bydra tot die bestuur van die water sisteem op die Potchefstroom Kampus van die Noord-Wes Universiteit.

Die doel van die studie is om 'n geïntegreerde geografiese databasis op te stel vir die drinkwater infrastruktuur op die Potchefstroom Kampus van die Noord Wes Universiteit. Die proefstudie fokus op geboue E4 en E6. ArcGIS 10 was gekies as die primêre sagteware sisteem wat aangewend kan word as medium om die problem voor te stel en op te los. ArcGIS se geografiese databasis dien as 'n houer waarin ruimtelike data in gestoor word. Data met betrekking tot drinkwater was ingesamel deur verskeie bronne te raadpleeg. Elektroniese en hardekopie CAD data was die algemene formaat vir die data.

'n Lêer geodatabasis was geskep in *ArcCatalog* met 'n standard koördinaat sisteem as verwysing vir die data in die geodatabasis. *ArcMap* is aangewend vir die 2D wysiging en die geografiese verwysing van die CAD tekening waarna 'n samestelling van die beskrywende data vir die komponente geskep is. Die eindresultaat van die datamodel en die data binne die geografiese databasis was om 'n 3D voorstelling in *ArcScene* voor te stel. Daardie voorstelling sal dus aangewend word vir 3D visualisering en die 3D analises van die data. Dit sal ook gemaklike toegang tot die beskrywende inligting en die verhoudings van die data bied. Verder sal die vermoë om 'n kortste roete analise uit te voer ook voorsien word.

Kernwoorde: GIS, geodatabase, water infrastruktuur, 3D, netwerk

# TABLE OF CONTENTS

## Chapter 1

### Introduction

1.1. Problem statement	3
1.2. Aim and objectives	4
1.3. Study area	4
1.4. Software overview	6
1.5. Dissertation outline	7

## Chapter 2

### Literature review

2.1. GIS, the system and the structure	8
2.2. Different types of data storage	11
2.2.1. Coverages	12
2.2.2. Shapefiles	15
2.2.3. Feature classes	16
2.2.4. Feature datasets	18
2.2.5. Geodatabases	18
2.2.5.1. Types of geodatabases	21
2.2.5.2. ArcSDE geodatabases	23
2.2.5.3. Enterprise GIS	24
2.3. Geodatabase design	26
2.3.1. Representations	26
2.3.2. Thematic layers	27
2.4. Inside a geodatabase, the structure and design	27
2.4.1. Important geodatabase design elements	28
2.4.1.1. Datasets	28
2.4.1.2. Relationship classes	29
2.4.1.3. Subtypes and domains	31
2.4.1.4. Topology	35
2.4.1.5. Geometric networks	42
2.4.1.6. Network datasets	49
2.5. CAD versus GIS	50
2.5.1. CAD	51
2.5.2. The differences between CAD and GIS	53
2.5.3. Problems of integrating CAD with GIS	54
2.5.4. Choosing CAD as main data source	55
2.6. Georeferencing CAD data	56

2.7. Co-ordinate systems	58
2.7.1. Geographical co-ordinate systems (GCS)	58
2.7.2. Spheroids, spheres and datums	59
2.7.3. Projected co-ordinate systems and projections	62
2.8. Case studies	65

### **Chapter 3**

#### Designing the geodatabase

##### The conceptual design

3.1. Database design	74
3.1.1. The ten geodatabase design steps	77
3.2. Applying the geodatabase design	78
3.2.1. The information products that will be produced with GIS	79
3.2.2. Identifying the key thematic layers	80

##### The logical design

3.3. The scale ranges and spatial representations for each thematic layer	81
3.4. Group representations into datasets	83
3.4.1. Feature datasets	83
3.4.2. Feature classes	83
3.5. Tabular database structure for attributes	84
3.5.1. Feature classes	84
3.5.2. Tables	92
3.5.3. The unique identifier	94
3.5.4. Subtypes	95
3.5.5. Domains	96
3.6. Define the spatial properties of the datasets	96
3.6.1. Topology rules	96
3.6.2. Applying spatial reference	99
3.7. Propose a geodatabase design	99
3.7.1. Relationship classes	99

### **Chapter 4**

#### Physical design concepts of the geodatabase

4.1. Implement, prototype, review and refine the design	103
4.1.1. Phase 1	103
4.1.2. Phase 2 – Importing CAD and digitizing in ArcMap	108
4.1.3. Phase 3 – The editing of line segments	118
4.2. Design work flows for building and maintaining each layer	121
4.2.1. Establishing relationship classes	124

4.2.2. Topology	132
4.2.3. Network dataset	133
4.2.4. Modelbuilder and the shortest route model	134
4.2.5. The fishnet	140
4.3. The final product	144
4.4. Documenting the design	148
<b>Chapter 5</b>	
Results	149
Conclusions of the results	154
<b>Chapter 6</b>	
Conclusions	156
Recommendations	158
<b>References</b>	159
<b>Addendums</b>	
Addendum A – Subtypes	
Addendum B – Domains	
Addendum C – The Geodatabase Diagram as as summary of the geodatabase	
Addendum D – A summary of the Geodatabase Diagram	

## LIST OF TABLES

Table 2.1. A compared summary of the three data storage types	17
Table 2.2. A summarized representation of the different databases	22
Table 2.3. The advantages of the file geodatabase in three sections	26
Table 2.4. The storage format for objects within the attribute table	29
Table 2.5. Split and merge policies	35
Table 2.6. A conceptual view of topology rules	39
Table 2.7. An analysis of geometric networks versus real life applications	45
Table 2.8. A summary of the edges versus junctions	46
Table 2.9. Simple and complex edges	47
Table 2.10. User-defined and orphan junctions	47
Table 2.11. A summary of the network datasets versus the geometric networks	49
Table 2.12. Comparison between the different projections	64
Table 3.1. Steps in database creation	77
Table 3.2. Ten steps to designing a geodatabase	78
Table 3.3. The scale ranges and the spatial representation for each layer	83
Table 3.4. PUK_Zones feature class	84
Table 3.5. PUK_Buildings features class	85
Table 3.6. Manhole_Chambers feature class	85
Table 3.7. PUK_Rooms feature class	86
Table 3.8. Thrust protection feature class	86
Table 3.9. Water mains feature class	87
Table 3.10. Geysers feature class	88
Table 3.11. Pumps feature class	88
Table 3.12. Control valve feature class	89
Table 3.13. Meters feature class	89
Table 3.14. End point facilities feature class	90
Table 3.15. System valve feature class	91
Table 3.16. Fittings features class	92
Table 3.17. List of contractors table	93
Table 3.18. Owner table	93
Table 3.19. Maintenance record	94
Table 3.20. A complete description of the “Unique_ID”	95

Table 3.21. A demonstration of the many-to-many relationship classes	101
Table 4.1. A representation of the different floor levels and the utilities based on those levels	119
Table 4.2. The creation of one-to-many relationships among the fields indicated	131
Table 5.1. Different tools and applications with regard to the model	149

## LIST OF FIGURES

Figure 1.1. The study area	6
Figure 2.1. Interactive maps of GIS	9
Figure 2.2. The transformation from paper maps	12
Figure 2.3. Coverages	13
Figure 2.4. Appearance of the coverage in an ArcCatalog directory	14
Figure 2.5. A detailed view of the coverage	14
Figure 2.6. Representation of the shapefiles in ArcCatalog	15
Figure 2.7. A geodatabase representation containing feature classes and feature datasets	18
Figure 2.8. Representation of the ArcSDE geodatabase	23
Figure 2.9. Enterprise GIS	25
Figure 2.10. Three types of relationship classes	30
Figure 2.11. Relationships between parcels	31
Figure 2.12. Subtypes of the “PressurizedMain”	32
Figure 2.13. Six new topology rules with ArGIS	36
Figure 2.14. Topology rules applied	37
Figure 2.15. The connection between different networks	43
Figure 2.16. A geometric network in its basic form	46
Figure 2.17. A layout of the geometric network	48
Figure 2.18. Representing CAD data as a number of transparent sheets	52
Figure 2.19. A representation of latitude and longitude lines	59
Figure 2.20. A representation of a geoid and an ellipsoid	60
Figure 2.21. The geoid	61
Figure 3.1. A simple entity relationship diagram	75
Figure 3.2. Topology rule defined between points and lines	97
Figure 3.3. Topology rule established for the polygons in the system	97
Figure 3.4. Topology rule established between lines and polygons	98
Figure 3.5. The relationship classes together with their cardinalities	100
Figure 3.6. The one-to-one relationship cardinality	101
Figure 3.7. The many-to-many relationship representation	101
Figure 3.8. A one-to-many relationship	101

Figure 4.1. The water feature dataset	104
Figure 4.2. The feature classes and relationships modelled in Microsoft Access	107
Figure 4.3. A design representation of the room polygons	109
Figure 4.4. The “Transformations” environment	111
Figure 4.5. The CAD drawing georeferenced over the QuikBird image (2008)	112
Figure 4.6. The layout of the PUK_Zones	114
Figure 4.7. The layout of the PUK_Buildings after the digitizing process	115
Figure 4.8. The PUK_Rooms after the digitizing process	116
Figure 4.9. A 3D representation of the digitized PUK_Rooms	117
Figure 4.10. Resembles the digitized PUK_Buildings in a 3D view	118
Figure 4.11. Attributes are easily located and revised in ArcScene	122
Figure 4.12. In the process of placing a point in ArcMap	123
Figure 4.13. The onset of a many-to-many relationship	125
Figure 4.14. The choice of a relationship cardinality	126
Figure 4.15. Two fields created in the many-to-many relationship	126
Figure 4.16. The many-to-many relationship table between main sections and fittings	127
Figure 4.17. The “Editor” toolbar with the “Attributes” tab	128
Figure 4.18. The attributes after opened by the “Attributes” tab in Figure 4.17	128
Figure 4.19. The “Identify” tool applied	129
Figure 4.20. A water main selected	130
Figure 4.21. Representation of the features participating in the topology rules	132
Figure 4.22. A representation of the topology rules applied	133
Figure 4.23. The shortest route model within the toolbox created	135
Figure 4.24. The fittings layer	137
Figure 4.25. The “first route” layer	137
Figure 4.26. The shortest route model is finished off	138
Figure 4.27. A representation of the shortest route between two points	138
Figure 4.28. A representation of the shortest route between more than two points	139
Figure 4.29. A representation of the extents of the QuickBird image (2008)	141
Figure 4.30. A representation of the cell sizes for the creation of the fishnet	143
Figure 4.31. A representation of the proposed buildings and the four levels in 3D	144
Figure 4.32. A representation of the rooms and facilities in 3D	145
Figure 4.33. A different angle of representation regarding Figure 4.32.	146
Figure 4.34. The final layout of buildings E4 and E6	147

Figure 5.1. The “Select By Attributes” tool applied	151
Figure 5.2. The results obtained with the “Select By Attributes” tool	151
Figure 5.3. The “Select By Location” tool applied	152
Figure 5.4. An indication of the five selected points along the water network	154
Figure 5.5. The shortest route application viewed from a different angle	154