# DEVELOPMENT OF A NATIONAL WATER CONSUMPTION ARCHIVE

Report to the
Water Research Commission

by

**JE van Zyl & L Geusteyn**

Department of Civil Engineering Science
University of Johannesburg

WRC Report No 1605/1/07
ISBN 978-1-77005-489-9

January 2007

## DISCLAIMER

This report has been reviewed by the Water Research Commission (WRC) and approved for publication. Approval does not signify that the contents necessarily reflect the views and policies of the WRC, nor does mention of trade names or commercial products constitute endorsement or recommendation for use

# EXECUTIVE SUMMARY

## Introduction

South Africa is a water-scarce country and has to constantly strive to apply its available water resources in the most efficient manner. Different users, including industry, domestic consumers, agriculture and the environment, vie for the available resources and have to be awarded an equitable and adequate share to ensure sustainability in the long term.

Domestic water consumption is a function of various factors, including stand size, income, household size, climate, pressure and water price. Although a number of studies have been done on various aspects of domestic water demand, we still have an inadequate understanding of how much water is consumed and how consumption is affected by the various influencing factors. South African design guidelines still use the same water demand estimates as originally developed in the early 1980s.

Some studies have indicated that the national guidelines overestimate average water demand for large stands, which translate into overly expensive networks and inefficient use of infrastructure. While certain municipalities have their own water demand guidelines, there is a strong need for a comprehensive database of consumption on a national level.

The overriding problem with studying water consumption is that quality data is difficult to get hold of. Municipal water meter readings are an ideal source of water demand information since they are taken on a regular basis by virtually all municipalities in the country. However, this information is generally difficult to access and are typically discarded (and thus lost) after a short period of time.

In recent years, GLS Consulting Civil Engineers developed a software product called Swift. This product allows access municipal treasury databases to obtain demographic and water consumption information of large numbers of users. Swift has been implemented at many local authorities throughout South Africa, covering different economic, socio-economic, climatic and other areas. It was seen as an imperative that action is taken to ensure that this data is not lost.

## Objectives

The main objective of the project was to develop a methodology for the archiving of water demand data, and to write an archive for the data currently available. This objective was achieved in full. The project also developed procedures for continuous expansion of the archive as new data becomes available.

## Methodology and Results

The study was lead by the Water Research Group at the University of Johannesburg in consultation and collaboration with GLS Consulting Engineers. It consists mainly of the collection, verification, cleaning and archiving of data on water demand and other user parameters. Software code and/or macros were developed to use on the existing and future databases. A non-propriety text storage format was selected and standardised. The procedures described above were then applied to the data in the existing Swift databases to archive the data.

Forty-eight municipal treasury databases were archived in this study. This includes four metros (Johannesburg, Tshwane, Ekurhuleni and Cape Town) and 151 cities or towns. The total number of stands in the databases exceeds 2.5 million, of which 1.5 million are non-vacant stands. The number of records (water meter records) in the databases exceed 2.7 million. In

most cases, the data for each non-vacant record includes actual water meter readings, reading dates and estimated monthly consumption figures for more than two years. Data for all types of users with metered consumption are included in the database, including domestic, commercial, industrial and educational users.

A method was developed and documented for the sustained archiving of future Swift databases. This will allow continued expansion of the water consumption archive with minimum effort after the completion of this project.

Finally, methods for making the data accessible to researchers and other interested parties in South Africa were investigated. Based on legal advice received, request for, and permission for the use of data by other interested parties for the sole purposes of research, shall be at the sole discretion of the WRC, subject to a written motivation and acceptance of an indemnity declaration (example in Appendix A).

**Conclusions**

A methodology to archive water demand data was developed, and the currently available data was written to the archive (approximately 1.5 million users). The parties involved in this project will continue to expand the archive as new demand data become available, although this falls outside the scope of this project.

The data archive will form a rich source for future research on water demand behaviours and trends. Aspects that can be studied include consumption of different user categories, variation of demand with geographical location and time, and the effect of changes in water price and pressure on demands.

## ACKNOWLEDGEMENTS

The research in this report emanated from a project funded by the Water Research Commission entitled *Development of a National Water Consumption Archive (*WRC Project K5/1605)

This project would not have been possible without financial support by the Water Research Commission. The authors would like to extend a word of appreciation for this opportunity. Mr JN Bhagwan, in particular, played a strong supporting and advisory role, which the authors gratefully acknowledge.

The authors would also like to acknowledge Kerry Fair, who assisted greatly with defining the archive structure and coding it.

# TABLE OF CONTENTS

# LIST OF FIGURES



# LIST OF TABLES

# CHAPTER 1: INTRODUCTION

## 1.1 BACKGROUND

South Africa is a water-scarce country and has to constantly strive to apply its available water resources in the most efficient manner. Different users, including industry, domestic consumers, agriculture and the environment, vie for the available resources and have to be awarded an equitable and adequate share to ensure sustainability in the long term.

Domestic water consumption is a function of various factors, including stand size, income, household size, climate, pressure and water price. Although a number of studies have been done on various aspects of domestic water demand, we still have an inadequate understanding of how much water is consumed and how consumption is affected by the various influencing factors. South African design guidelines still use the same water demand estimates as originally developed in the early 1980s.

Some studies have indicated that the national guidelines overestimate average water demand for large stands, which translate into overly expensive networks and inefficient use of infrastructure. While certain municipalities have their own water demand guidelines, there is a strong need for a comprehensive database of consumption on a national level.

The overriding problem with studying water consumption is that quality data is difficult to get hold of. Even a large logging exercise can realistically only aim to reach a small proportion of households. Municipal water meter readings are an ideal source of water demand information since they are taken on a regular basis by virtually all municipalities in the country. However, this information is generally difficult to access and are typically discarded (and thus lost) after a short period of time.

In recent years, GLS Consulting Civil Engineers developed a unique software product called Swift. This product allows access municipal treasury databases to obtain demographic and water consumption information of large numbers of users. Swift has been implemented at many local authorities throughout South Africa, covering different economic, socio-economic, climatic and other areas.

Examples of cities and towns that have implemented Swift include Tshwane Metropolitan Municipality, Ekurhuleni, Johannesburg and most municipalities in the Western Cape. Data is verified through associated bulk water meter readings where possible. The existing databases cover years of consumption data for millions of consumers. There can hardly be a more valuable source of information on the topic of water consumption both now and in the future. It was seen as an imperative that action is taken to ensure that this data is not lost. This project aimed to collect existing and future water demand data in an archive that will be accessible to researchers and other interested parties in South Africa.

## 1.2 OBJECTIVES AND DELIVERABLES

The following objectives were set for the project:

- To identify water consumption data to be included in the consumption data archive from Swift and other sources.
- To develop a methodology for abstracting and storing the water consumption data.
- To archive the water consumption data currently available from Swift databases of different municipalities in South Africa.
- To develop a procedure for sustained archiving of future water consumption data as it

becomes available.

## 1.3 METHODOLOGY

The study was lead by the Water Research Group at the University of Johannesburg in consultation and collaboration with GLS Consulting Engineers. It consists mainly of the collection, verification, cleaning and archiving of data on water demand and other user parameters.

Permission was obtained from the relevant local authorities to use their data for research purposes. The Swift databases include a large array of data on different user parameters. These were reviewed to determine the relevant data fields to include in the archive. Issues such as the protection of privacy were considered and steps were taken to ensure that individual rights are respected and protected.

Software code and/or macros were developed to use on the existing and future databases. A generally accessible storage format was selected and standardised. The procedures described above were then applied to the data in the existing Swift databases to archive the data.

A method was developed and documented for the sustained archiving of future Swift databases. This will allow for the continued expansion of the water consumption archive with minimum effort after the completion of this project. Finally, methods for making the data accessible to researchers and other interested parties in South Africa were investigated.

# CHAPTER 2: DATA

## 2.1 INTRODUCTION

In recent years, GLS Consulting Civil Engineers developed a software product called Swift. This product allows them to access municipal treasury databases to obtain demographic and water consumption information on large numbers of users. Swift has been implemented at many local authorities throughout South Africa, covering different economic, socio-economic, climatic and other areas.

The most recent data collected with Swift is used for various management and engineering functions. However, older data is of little use and is often discarded. The main purpose of this project was to preserve the data in a generic and easily accessible form to ensure that it is not lost. The data can be used with much benefit in future to study water demand behaviour and geographical and temporal trends in water consumption. It can form the basis of future studies to investigate the effect of changes, such as the introduction of pressure management, increases in water tariffs and densification in a certain area on water consumption patterns.

## 2.2 SWIFT PROGRAM

It is important to understand the purpose and structure of the Swift programme to understand how data is obtained, manipulated and stored.

### 2.2.1 Purpose of the program

In the quest for a water service provider to optimise the service to consumers in terms of providing sufficient water at an affordable cost, the need for accurate information about the system is becoming ever more important. This information is in fact available in Municipal Treasury Databases where the readings for all water meters in the system are stored on a monthly basis along with land use and other stand related information. The problem is that the treasury systems were not designed to produce the information and statistical reports required by the infrastructure managers. The data is also not spatially referenced, which would allow links to cadastral database applications in GIS. It is therefore necessary to extract the information from the databases and use additional manipulation through software designed for the purpose, to perform the required statistical calculations.

Although the calculations needed to produce the information are theoretically simple, the volume of data that is involved makes it difficult to use spreadsheets. Even in relatively small towns, where there are less than 15 000 stands, it has been found that spreadsheets become prohibitively inefficient and clumsy.

Swift was designed specifically for the purposes of the infrastructure manager and as such produces, from large quantities of information, consumption related reports required, if not by law, then for the effective management of the system. Swift has been used to produce statistics used in:

- Water demand management initiatives.
- Water audits.
- Water loss calculations.
- Calculation of water tariffs.
- Water consumptions profiles for user defined categories (e.g. land use).
- Water and sewer master plans.

- Identification of faulty meter readings.
- Population of water and sewer model databases.
- In addition a user-friendly query builder allows the system manager to perform any further queries on the data that may be required.

### 2.2.2 Link to the Treasury System

The Swift databases are obtained from the relevant authorities' treasury databases. As such Swift includes data for every stand and every meter in the study area. Fields in the databases include information on stand related data such as the owner, consumer, address, land use, zoning, consumption and tax tariffs of the stand. The value of the stand and any improvements are also included. Information related to the meters includes the meter readings, the meter serial numbers and the date of installation.

During the course of various studies it has been found that it is important to validate the historical meter records extracted from the treasury systems. Routines are required to identify and correct irregularities in the water meter readings such as sudden drops in the readings when a meter is replaced, spikes and/or dips in the water consumption records and unrealistically high or low readings. Possible causes for these errors include both faulty water meters and data capturing errors.

There are also basic functions to check the integrity of the latest readings against the historical meter records.

Swift allows the user to customize certain settings so that the analyses can be performed for any treasury database. For example, what might be considered a large consumer in one town would not necessarily be a large user in another, larger town. Swift allows the user to define the consumption that determines the large users. Similarly, the suburbs and land use information can be categorized into user specified groups so that meaningful results can be obtained for any system.

### 2.2.3 Data fields

The typical data fields that form part of a Swift project are listed in Table 1. The fields included in the archive are also identified.

**Table 1: Typical data fields included in a Swift project**

| Field Name | Type | Description | Included in archive |
|---|---|---|---|
| Stand_ID | String | Stand's ID consisting of Town, Suburb, Stand, Portion and Sub portion. The format is defined according to the treasury system. | Y |
| Unit_No | String | Additional Stand ID number to make record unique | Y |
| NetworkNr | String | IMQS Network Number | N |
| WardNr | String | Ward Number (Political Classification) | Y |
| Treasury_ID | String | Unique identification number carried by the treasury database | Y |
| Meter_Count | Integer | Count of records on the stand | N |
| Units_T | Integer | Number of living units connected to the record | Y |
| Units | Integer | Calculated Number of living units on the stand | Y |
| Units_LU | Integer | Theoretical number of units. If Units_ha (according to stand's landuse category and suburb category) > 0 then the value = Units_ha multiplied by the stand area, otherwise if Unit_ha = 0 the value = Units. | N |
| Units_Z | Integer | Theoretical number of units. If Units_ha (according to stand's zoning category and suburb category) > 0 then the value = Units_ha multiplied by the stand area, otherwise if Unit_ha = 0 the value = Units. | N |

| | | | |
|---|---|---|---|
| Account_No | String | Account number | Y |
| Stand_Owner | String | Stand owner | Y |
| Address | String | Address | Y |
| GIS_Code | String | GIS Code (as supplied by the treasury) | Y |
| GIS_Adj | String | GIS Code after adjustment | |
| Consumer | String | Consumer | Y |
| Metered | True/ False | Whether or not the water consumption is metered | N |
| Meter_ Installation | Date | Date when meter was installed | Y |
| Serial_No | String | Meter serial number | Y |
| Route_No | String | Meter reader route number | Y |
| Measurement_ Unit | String | Water meter measurement unit | Y |
| Display_Size | Integer | The number of digits on the meter display | Y |
| Tax_Tariff | String | Tax tariff code | Y |
| Landuse | String | Landuse code | Y |
| Zoning | String | Zoning code | Y |
| Avail_Code | String | Water availability code | Y |
| Water_Tariff | String | Water consumption tariff code | Y |
| Value_Site | Integer | Value of the site | Y |
| Value_improvement | Integer | Value of the improvements on the site | Y |
| Stand_Area | Integer | Area of the stand | Y |
| Area_Unit | String | Unit of the area value | Y |
| Area_Division | Integer | The stand is categorised according to its area and the area category definitions as defined by the user | N |
| Town | String | Town | Y |
| Suburb | String | Suburb | Y |
| Suburb_Category | String | The user specified suburb category to which the suburb code is mapped | Y |
| Stand_No | String | Stand number | Y |
| Portion | String | Portion | Y |
| Large | True/ False | If the stand AADD > large AADD definition the large = true. If the land use category is a multiple unit category large = stand AADD/units > large | N |
| Vacant | True/ False | If all the records for the stand are vacant (AADD < 0.1 kl/day) the value is true | N |
| Node | Integer | The Wadiso node nearest to the stand | N |
| Manhole | String | The sewer manhole nearest to the stand | N |
| Land_Category | String | The land use category determined from the land use, water tariff, availability code and tax tariff code | Y |
| Zone_Category | String | The zoning category determined from the zoning code and the land use category | Y |
| Av_Day_Demand | Number | Annual average daily demand for the record determined from the readings | Y |
| AADD_Units | Number | Av_day_demand divided by the number of units supplied by the meter (Units_T) | N |
| Stand_AADD | Number | The sum of the Av_day_demands for all the records on the stand | N |
| Total_Measurement | Number | The total water consumption in the last year. If the record period is less than a year the total water demand = Av_Day_Demand x 365 | Y |
| Adjust | Number | Adjustment code indicating anomalies in the readings | Y |
| Ratio1 | Number | If Ave_Day_demand >= Ratio1 Cut-off: Ratio1 = average daily demand of the last month over the annual average daily demand. If Ave_Day_demand < Ratio1 Cut-off: Ratio1 = 0 | N |
| Ratio3 | Number | If Ave_Day_demand >= Ratio3 Cut-off: Ratio3 = average daily demand of the last 3 months over the annual average daily demand. If Ave_Day_demand < Ratio3 Cut-off: Ratio3 = 0 | N |
| Unit_AADD_LU | Number | Theoretical unit water demand (kl/day) according to the land use category and the suburb category as defined in the theoretical values table. If the Unit_Ha for the suburb category and the land use category > 0 then the value represents the theoretical demand per unit. Otherwise the value represents the demand for | N |

| | | | |
|---|---|---|---|
| | | the stand. | |
| Unit_AADD_Z | Number | Theoretical unit water demand (kl/day) according to the zoning category and the suburb category as defined in the theoretical values table. If the Unit_Ha for the suburb category and the zoning category > 0 then the value represents the theoretical demand per unit. Otherwise the value represents the demand for the stand. | N |
| UAW_Act | Number | Present theoretical water losses as defined for the suburb category in the suburb categories table | N |
| UAW_Theor | Number | Future theoretical water losses as defined for the suburb category in the suburb categories table | N |
| UH_LU | Number | Unit hydrograph according to the land use category and the suburb category as defined in the theoretical values table. | N |
| UH_Z | Number | Unit hydrograph according to the Zoning category and the suburb category as defined in the theoretical values table. | N |
| Tariff | Integer | Tariff Category according to the Water Tariff code or Avail Code as defined in the tariff categories table | Y |
| Profile | Integer | Tariff category according to the Landuse Category as defined in the profile categories table | N |
| Avail_Tariff | Integer | Tariff Category according to the Avail Code as defined in the tariff categories table | N |
| Non_Pay | Number | Percentage of non payment according to the tariff and suburb category as defined in the Theoretical Tariff values table | N |
| Growth | Number | Percentage of growth according to the tariff and suburb category as defined in the Theoretical Tariff values table | N |
| Measurement_1 | Integer | Latest meter reading | N |
| Measurement_Date_1 | Date | Latest meter reading date | N |
| Measurement_Type_1 | String | Latest meter reading type | N |
| NumReadings | 25 | Number of readings in the record | N |
| Readings | Binary | The readings | Y |
| NumOriginal | Integer | Number of readings in the stream which has been corrected | Y |
| Original | Binary | The original readings where corrections have been made by the user | Y |
| OrigAdj | String | The original adjustment code before the user corrected the readings | Y |
| NumNewReadings | Integer | Number of readings imported for the last month | N |
| NewReadings | Binary | Last month's readings | N |
| NewWarnCode | Integer | Adjustment codes after latest month's readings have been included | N |
| Readings | Binary | The readings | Y |

## 2.3 DESCRIPTION OF DATA

Treasury data for a number of municipalities were collected in January 2005 from various Swift databases. Tables 2 to 4 indicate what data was successfully collected and extracted. This forms the basis of the archive, although steps will be taken to continuously expand the archive as new data becomes available. The extracted data represents the monthly consumption records of almost three million users for periods ranging from twelve to sixty months.

Forty-eight municipal treasury databases were collected and extracted for archiving in this study. This includes four metros (Johannesburg, Tshwane, Ekurhuleni and Cape Town) and 151 cities or towns. The total number of stands in the databases exceeds 2.5 million, of which 1.5 million are non-vacant stands. The number of records (water meter records) in the databases exceed 2.7 million. In most cases, the data for each non-vacant record includes actual water meter readings, reading dates and estimated monthly consumption figures for more than two years. Data for all types of users with metered consumption are included in the database, including domestic, commercial, industrial and educational users.

Stands were classified as ‘vacant’ if the Vacant record in the municipal treasury database was set to true for the stand. However, not all of these ‘vacant’ stands had zero water consumption;

quite a number had relevant and regular readings. It may thus be that these stands had been vacant but had since been occupied but the treasury systems had not been updated.

Table 2 lists the municipalities and associated towns and gives the main characteristics of each area, including its development characteristic (rural/urban), the towns or suburbs included in it, an economic and development attribute (metro/city/town, coastal/inland) and its water region according to the Department of Water Affairs and Forestry (DWAF) classification system.

Table 3 lists the municipalities and their user data characteristics, including the number of records and stands, number of vacant records and stands, number of occupied stands, large user definition, number of large user stands and meters, number and fraction of domestic stands, fraction of records with zero stand area and value, and fraction of stands with unknown land use.

Table 4 lists the municipalities and their consumption data characteristics, including the number of records, total number of stands, number of vacant records, number of monthly readings, period of readings, total water demand per year, average annual daily demand (AADD), unit demand per stand, and unit demand per meter.

**Table 2: Demand area characteristics**

| Municipality | Towns and Cities | Period of Readings | # of Stands |
|---|---|---|---|
| Beaufort West | Beaufort West | Dec 02 - Feb 04 | 6,941 |
| BergRiver | Velddrif, Porterville, Piketberg, Laaiplek, Aurora, Eendekuil, Dwarskersbos, Redelingshuys | Feb 03 - Mar 04 | 10,012 |
| Blaauwberg | Blaauwberg, Milnerton, Mamre, Atlantis, Melkbos | Jun 00 - Jul 02 | 49,298 |
| Breede River | Robertson, Bonnievale, Ashton, Montagu, McGregor | Oct 02 - Oct 04 | 20,079 |
| Breede Valley | Worcester | Nov 99 - Jan 04 | 15,075 |
| Buffalo City | East London | Jan 03 - Feb 05 | 119,748 |
| Cape Agulhas | Bredasdorp, Waenhuiskrans, Struisbaai, Suiderstrand, Napier, Klipdale, Lagulhas | Mar 99 - Feb 04 | 8,576 |
| Cederberg | Clanwilliam, Citrusdal, Lamberts Bay, Graafwater, Elands Bay, Leipoldtville | Dec 02 - Dec 03 | 6,881 |
| Drakenstein | Wellington, Paarl, Gouda, Bainskloof, Hermon, Mbekweni | Dec01 - Dec 03 | 37,076 |
| Ekurhuleni | Alberton | Aug 02 - Sep 04 | 40,065 |
| Ekurhuleni | Germiston | Aug 02 - Sep 04 | 117,846 |
| Ekurhuleni | Kempton Park | Aug 02 - Sep 04 | 87,779 |
| Ekurhuleni | Lethabong | Aug 02 - Sep 04 | 25,150 |
| Ekurhuleni | Nigel | Sep 02 - Oct 04 | 27,420 |
| Ekurhuleni | Springs | Sep 02 - Oct 04 | 45,461 |
| Ekurhuleni | Benoni | Sep 02 - Oct 04 | 100,921 |
| Ekurhuleni | Boksburg | Aug 02 - Sep 04 | 72,334 |
| Ekurhuleni | Brakpan | Sep 02 - Oct 04 | 63,812 |
| George | George, Heroldsbay, Wildernis, Twee Rivieren | Jul 99 - Jul 03 | 33,229 |
| Helderberg | Gordons Bay, Strand, Somerset West, Maccasar | Apr 98 - Apr 03 | 37,830 |
| Johannesburg | Lethabong | Feb 02 - Mar 04 | 6,398 |
| Johannesburg | Midrand | Feb 02 - Mar 04 | 39,699 |
| Johannesburg | Northern Johannesburg (Sandton, Randburg,Bryanston, Sunninghill, Morningside) | Jan 02 - Feb 04 | 123,262 |
| Johannesburg | Southern Johannesburg 1 | Feb 02 - Mar 04 | 96,235 |
| Johannesburg | Southern Johannesburg (Soweto) | Feb 02 - Mar 04 | 160,076 |

| Municipality | Towns and Cities | Period of Readings | # of Stands |
|---|---|---|---|
| Johannesburg | Eastern Johannesburg | Feb 02 -Mar 04 | 118,315 |
| Johannesburg | Western Johannesburg | Feb 02-Mar 04 | 105,494 |
| Langeberg | Albertina, Gouritsmond, Heidelberg, Jongensfontein, Riversdale, Slangrivier, Stilbaai, Theronsville, Witsand | Dec 02 -Jan 04 | 16,008 |
| Matzikama | Klawer, Lutzville, Vanrhynsdorp, Vredendal | Jan 02 - Feb 04 | 8,317 |
| Mossel Bay | Mossel Bay, Dana Bay, Klein Brak, Hartenbos | Oct 01 - Oct 03 | 32,157 |
| Oostenberg | Brackenfell, Blue Downs, Kuilsrivier | Nov 99 - Dec 01 | 61,300 |
| Oudtshoorn | Oudtshoorn | Nov 01 - Nov 03 | 14,335 |
| Overstrand | Gansbaai, Hermanus, Kleinmond, Stanford | Feb 00 -Feb 04 | 30,307 |
| Plettenberg Bay | Plettenberg Bay | Dec 99 - Nov 04 | 9,155 |
| Randfontein | Randfontein, Mohlakeng | Apr 01 - May 03 | 17,979 |
| Saldanha Bay | Langebaan, Paternoster, Jacobsbaai, Vredenburg, St Helena Bay, Hopefield, Ongegund, Louwville | Apr 99 - Mar 04 | 23,481 |
| Sedibeng | Sharpville, Sebokeng, Boipatong, Evaton, Stretford, Vereeniging, Meyerton, Kanana, Hartebeestfontein, Klipwater, Walkerville | June 02 - Jul 04 | 170,126 |
| Stellenbosch | Stellenbosch | Oct 99 - Oct 01 | 14,335 |
| Swartland | Abbotsdale, Chatsworth, Grottobaai, Darling, Malmesbury, Morreesburg, Koringberg, Riebeeck Wes, Riebeeck Kstl, Yzerfontein | Oct 01 - Oct 03 | 22,763 |
| Theewatersklf | Caledon, Grabouw, Genadedal, Greyton, Riversonderend, Villiersdorp | Sep 03 -Oct 04 | 16,305 |
| Tshwane | Atteridgeville | Feb 02 - Feb 05 | 17,753 |
| Tshwane | Centurion | Dec 02 -Jan 05 | 35,652 |
| Tshwane | Mamelodi | Feb 02 - Feb 05 | 38,493 |
| Tshwane | NPMSS - Akasia, Rosslyn, Shoshanghuve | Feb 03 -Mar 05 | 84,900 |
| Tshwane | Odi (Garankua) | Dec 02 - Mar 05 | 45,793 |
| Tshwane | Pretoria | Feb 02 -Feb 05 | 146,691 |
| Tshwane | Temba (Hammanskraal) | Jan 04 - Feb 05 | 12,108 |
| Tygerberg | Bellville, Cape Town, Delft, Durbanville, Goodwood, Mfuleni, Khayelitsha, Parow | Aug 01 - Aug 03 | 196,036 |
| **TOTAL** | | | **2,589,006** |

**Table 3: User data characteristics**

| Municipality | Number of records | Total Number of Stands | Number of Vacant records | Number of Vacant Stands | Number of Occupied Stands | Large User Definition (kl/day) | Number of Large Users (Stands) | Number of Large Users (meters) | Number of domestic stands | % domestic stands of all stands | % of all records with zero stand area | % of all records with zero stand value | % of all stands with unknown land use assigned |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Beaufort West | 7,443 | 6,941 | 878 | 830 | 6,111 | 20 | 13 | 32 | 6,227 | 89.71% | 31.66% | 26.39% | 6.21% |
| BergRiver | 10,012 | | | | 3,626 | 20 | | 5,258 | | | 10.03% | 17.36% | 37.09% |
| Blaauwberg | 55,362 | 49,298 | 20,509 | 16,415 | 32,883 | 50 | 51 | 59 | 40,815 | 82.73% | 0.37% | 18.14% | 10.75% |
| Breede River | 20,612 | 20,079 | 9,249 | 9,244 | 10,835 | 20 | 7,064 | 4,405 | 7,408 | 36.89% | 25.21% | 44.10% | 0.03% |
| Breede Valley | 15,389 | 15,075 | 2,299 | 2,285 | 12,790 | 20 | 58 | 71 | 13,902 | 92.22% | 4.78% | 2.48% | 0.03% |
| Buffalo City | 129,453 | 119,748 | 48,030 | 47,877 | 71,871 | 20 | 234 | 1,498 | 102,665 | 85.73% | 0.17% | 11.72% | 9.85% |
| Cape Agulhas | 8,588 | 8,576 | 1,869 | 1,867 | 6,709 | 20 | 5 | 5 | 8,136 | 94.87% | 1.79% | 6.25% | 0.00% |
| Cederberg | 7,158 | 6,881 | 2,308 | 2,307 | 4,574 | 20 | 39 | 96 | 6,838 | 99.38% | 3.20% | 18.51% | 0.00% |
| Drakenstein | 40,772 | 37,076 | 11,110 | 11,046 | 26,030 | 50 | 297 | 2,090 | 15,647 | 42.20% | 0.11% | 11.27% | 30.08% |
| Ekurhuleni 1 | 41,365 | 40,065 | 6,925 | 6,879 | 33,186 | 20 | 151 | 541 | 34,055 | 85.00% | 0.00% | 1.03% | 2.85% |
| Ekurhuleni 2 | 122,569 | 117,846 | 42,378 | 42,094 | 75,752 | 20 | 420 | 1,322 | 107,602 | 91.31% | 0.37% | 14.37% | 2.42% |
| Ekurhuleni 3 | 90,163 | 87,779 | 33,753 | 33,532 | 54,247 | 20 | 233 | 619 | 82,262 | 93.71% | 0.13% | 20.73% | 0.36% |
| Ekurhuleni 4 | 25,843 | 25,150 | 9,374 | 9,345 | 15,805 | 20 | 90 | 169 | 19,217 | 76.41% | 13.61% | 20.50% | 16.74% |
| Ekurhuleni 5 | 29,662 | 27,420 | 6,435 | 5,829 | 21,591 | 20 | 42 | 494 | 24,142 | 88.05% | 3.24% | 12.54% | 1.88% |
| Ekurhuleni 6 | 46,884 | 45,461 | 12,497 | 12,461 | 33,000 | 20 | 85 | 259 | 40,267 | 88.57% | 0.00% | 4.82% | 1.92% |
| Ekurhuleni 7 | 104,407 | 100,921 | 46,503 | 46,448 | 54,473 | 20 | 141 | 474 | 95,160 | 94.29% | 0.00% | 0.26% | 0.36% |
| Ekurhuleni 8 | 82,503 | 72,334 | 22,267 | 14,040 | 58,294 | 20 | 194 | 784 | 65,672 | 90.79% | 0.00% | 20.90% | 1.62% |
| Ekurhuleni 9 | 64,776 | 63,812 | 49,400 | 49,327 | 14,485 | 20 | 56 | 220 | 60,988 | 95.57% | 0.01% | 8.43% | 0.51% |
| George | 33,292 | 33,229 | 8,155 | 8,151 | 25,078 | 20 | | 48 | 13,596 | 54.09% | 3.15% | 4.11% | 41.66% |
| Helderberg | 38,420 | 37,830 | 4,755 | 4,656 | 33,174 | 50 | 26 | 44 | 31,948 | 84.45% | 9.41% | 9.34% | 11.04% |
| Johannesburg 1 | 6,594 | 6,398 | 2,923 | 2,914 | 3,484 | 20 | 3 | 84 | 2,889 | 45.15% | 34.05% | 46.07% | 50.23% |
| Johannesburg 2 | 41,086 | 39,699 | 23,134 | 23,044 | 16,655 | 20 | 91 | 224 | 31,593 | 79.58% | 0.05% | 7.69% | 3.81% |
| Johannesburg 3 | 132,061 | 123,262 | 67,218 | 64,847 | 58,415 | 20 | 490 | 1,736 | 89,714 | 72.78% | 12.40% | 17.06% | 12.94% |
| Johannesburg 4 | 106,225 | 96,235 | 48,743 | 46,198 | 50,037 | 20 | 528 | 3,633 | 65,375 | 67.93% | 0.08% | 24.02% | 22.34% |
| Johannesburg 5 | 165,967 | 160,076 | 149,650 | 146,682 | 13,394 | 20 | 154 | 1,005 | 136,348 | 85.18% | 0.86% | 19.44% | 11.47% |
| Johannesburg 6 | 128,483 | 118,315 | 53,668 | 52,324 | 65,991 | 20 | 1,376 | 3,276 | 95,515 | 80.73% | 7.14% | 29.11% | 8.80% |

| | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Johannesburg 7 | 118,793 | 105,494 | 78,386 | 69,016 | 36,478 | 20 | 258 | 1,448 | 89,696 | 85.02% | 0.05% | 15.24% | 9.47% |
| Langeberg | 18,682 | 16,008 | 7,091 | 6,999 | 9,009 | 10 | 154 | 1,565 | 15,170 | 94.77% | 0.08% | 35.00% | 1.21% |
| Matzikama | 8,706 | 8,317 | 1,637 | 1,545 | 6,772 | 10 | 41 | 103 | 7,833 | 94.18% | 35.55% | 21.84% | 0.00% |
| Mossel Bay | 34,439 | 32,157 | 15,963 | 15,804 | 16,353 | 20 | 45 | 385 | 26,387 | 82.06% | 17.54% | 29.26% | 0.43% |
| Oostenberg | 62,076 | 61,300 | 5,461 | 4,923 | 56,377 | 20 | 99 | 104 | 57,849 | 94.37% | 25.00% | 12.75% | 2.36% |
| Oudtshoorn | 23,366 | | 12,000 | | | | | | | | 42.58% | 46.52% | 100.00% |
| Overstrand | 30,515 | 30,307 | 9,236 | 9,185 | 21,122 | 20 | 28 | 32 | 21,805 | 71.95% | 4.57% | 7.60% | 17.79% |
| Plettenberg Bay | 9,299 | 9,155 | 1,696 | 1,688 | 7,467 | 10 | 39 | 112 | 7,305 | 79.79% | 7.23% | 6.10% | 14.08% |
| Randfontein | 19,304 | 17,979 | 174 | 139 | 17,840 | 15 | 47 | 238 | 16,055 | 89.30% | 1.93% | 1.85% | 5.26% |
| Saldanha Bay | 23,750 | 23,481 | 6,589 | 6,565 | 16,916 | 20 | 73 | 119 | 21,564 | 91.84% | 2.05% | 10.08% | 1.25% |
| Sedibeng | 172,383 | 170,126 | 129,554 | 129,357 | 40,769 | 20 | 213 | 773 | 144,135 | 84.72% | 0.95% | 4.88% | 4.75% |
| Stellenbosch | 15,250 | 14,335 | 3,249 | 3,162 | 11,173 | 20 | 98 | 131 | 11,551 | 80.58% | 4.38% | 8.21% | 16.13% |
| Swartland | 23,652 | 22,763 | 9,018 | 9,014 | 13,749 | 10 | 62 | 382 | 17,038 | 74.85% | 19.46% | 16.21% | 10.92% |
| Theewaterskloof | 16,852 | 16,305 | 6,230 | 6,180 | 10,125 | 20 | 19 | 124 | 10,186 | 62.47% | 9.74% | 12.84% | 33.58% |
| Tshwane 1 | 18,063 | 17,753 | 2,899 | 2,839 | 14,914 | 20 | 30 | 98 | 17,257 | 97.21% | 12.77% | 12.75% | 1.97% |
| Tshwane 2 | 37,253 | 35,652 | 1,742 | 1,703 | 33,949 | 20 | 139 | 318 | 30,696 | 86.10% | 6.07% | 5.98% | 5.04% |
| Tshwane 3 | 7,557 | 5,737 | 7,552 | 5,733 | 4 | 20 | 2 | 2 | 1,399 | 24.39% | 68.48% | 68.52% | 75.09% |
| Tshwane 4 | 38,642 | 38,493 | 7,184 | 7,168 | 31,325 | 20 | 48 | 78 | 31,151 | 80.93% | 12.14% | 12.18% | 18.63% |
| Tshwane 5 | 88,102 | 84,900 | 19,958 | 19,814 | 65,086 | 20 | 131 | 319 | 79,925 | 94.14% | 0.00% | 8.25% | 18.36% |
| Tshwane 6 | 59,171 | 45,793 | 1,263 | 1,217 | 44,576 | 20 | 103 | 2,516 | 45,023 | 98.32% | 100.00% | 0.00% | 0.02% |
| Tshwane 7 | 174,000 | 146,691 | 43,546 | 38,917 | 107,774 | 20 | 974 | 3,703 | 105,101 | 71.65% | 20.21% | 20.41% | 25.37% |
| Tshwane 8 | 12,231 | 12,108 | 933 | 929 | 11,179 | 20 | 42 | 111 | 11,754 | 97.08% | 100.00% | 100.00% | 0.00% |
| Tygerberg | 224,878 | 196,036 | 71,139 | 62,385 | 133,651 | 20 | 423 | 4,362 | 185,093 | 94.42% | 8.35% | 23.45% | 0.43% |
| **TOTAL** | **2,792,053** | **2,570,396** | **1,126,530** | **1,064,924** | **1,509,098** | **1,005** | **14,909** | **45,469** | **2,151,954** | **83.72%** | **8.42%** | **15.89%** | **9.13%** |

**Table 4: Consumption data characteristics**

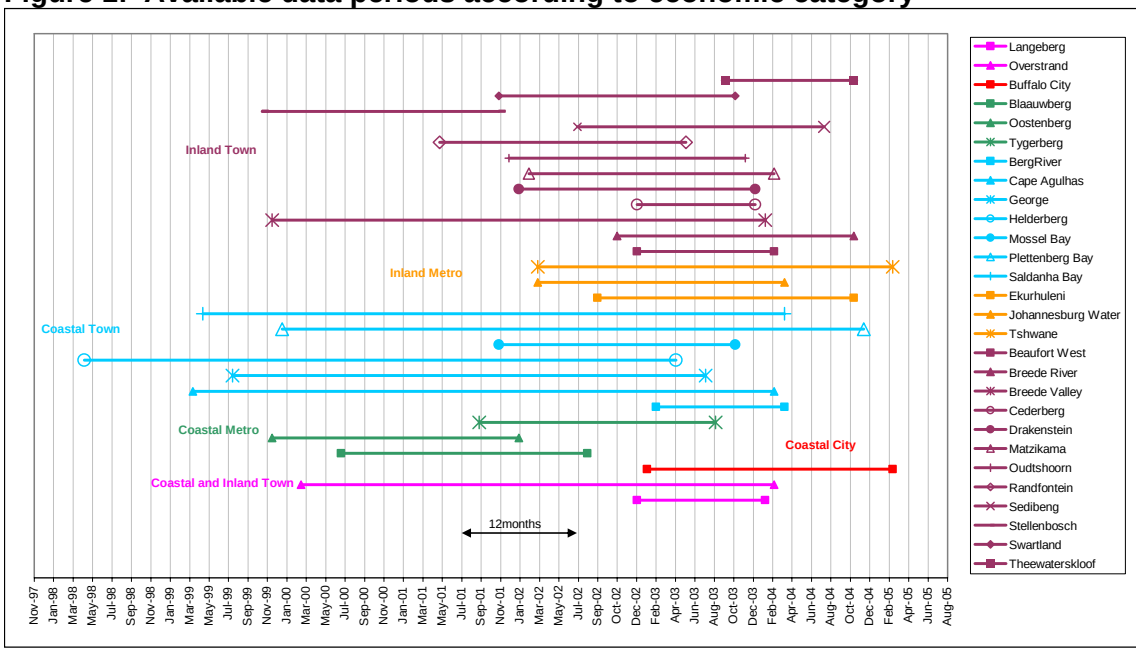
| Municipality | Number of records | Total Number of Stands | Number of Vacant records | Number of monthly readings | Period of Readings (more or less) | Total Water Demand per Year (kL/yr) | AADD (kL/Day) | Unit Demand per Stand (kL/Day) | Unit Demand per Meter (kL/Day) |
|---|---|---|---|---|---|---|---|---|---|
| Langeberg | 18,682 | 16,008 | 7,091 | 25 (majority) range 1 -42 | Dec02 -Jan 04 | 4316426 | 12453 | 1.38 | 1.07 |
| Overstrand | 30,515 | 30,307 | 9,236 | 48 (majority) | Feb 00 -Feb 04 | 5382502 | 14747 | 0.70 | 0.69 |
| Buffalo City | 129,453 | 119,748 | 48,030 | 25 | Jan 03 - Feb 05 | 25466982 | 73713 | 1.03 | 0.91 |
| Blaauwberg | 55,362 | 49,298 | 20,509 | 24 | Jun 00 - Jul 02 | 17658818 | 51789 | 1.05 | 0.94 |
| Oostenberg | 62,076 | 61,300 | 5,461 | 25 | Nov 99 – Dec 01 | 18026358 | 54577 | 0.97 | 0.97 |
| Tygerberg | 224,878 | 196,036 | 71,139 | 24 | Jul/Aug 01 - Jul/Aug 03 | 50582972 | 143210 | 1.07 | 0.94 |
| BergRiver | 10,012 | | | 13 (majority) range (1 -13) | Feb 03 - Mar 04 | 2018794 | 5531 | | 1.78 |
| Cape Agulhas | 8,588 | 8,576 | 1,869 | 60 (majority) range of 1- 60 | Mar 99 - Feb 04 | 1494697 | 4143 | 0.62 | 0.62 |
| George | 33,292 | 33,229 | 8,155 | 48 (majority( range 12 -48) | Jul 99 - Jul 03 | 5852239 | 16817 | 0.67 | 0.67 |
| Helderberg | 38,420 | 37,830 | 4,755 | 60 | Apr 98 - Apr 03 | 12616328 | 35164 | 1.06 | 1.05 |
| Mossel Bay | 34,439 | 32,157 | 15,963 | 24 | Oct 01 - Oct 03 | 5219693 | 14801 | 0.91 | 0.58 |
| Plettenberg Bay | 9,299 | 9,155 | 1,696 | 60 (majority) range of 1-60 | Dec 99 - Nov 04 | 2015392 | 5537 | 0.74 | 0.73 |
| Saldanha Bay | 23,750 | 23,481 | 6,589 | 60 (majority) range of 1-60 | Apr 99 - Mar 04 | 7917047 | 22601 | 1.34 | 1.32 |
| Ekurhuleni 1 | 41,365 | 40,065 | 6,925 | 25 (majority) range 1 -36 | Aug 02 - Sep 04 | 18378994 | 56392 | 1.70 | 1.64 |
| Ekurhuleni 2 | 122,569 | 117,846 | 42,378 | 25 (majority) | Aug 02 - Sep 04 | 39374500 | 112446 | 1.48 | 1.40 |
| Ekurhuleni 3 | 90,163 | 87,779 | 33,753 | 25 (majority) | Aug 02 - Sep 04 | 24072152 | 68561 | 1.26 | 1.22 |
| Ekurhuleni 4 | 25,843 | 25,150 | 9,374 | 25 (majority) | Aug 02 - Sep 04 | 8228412 | 23549 | 1.49 | 1.43 |
| Ekurhuleni 5 | 29,662 | 27,420 | 6,435 | 25 (majority) | Sep 02 - Oct 04 | 9500665 | 26387 | 1.22 | 1.14 |
| Ekurhuleni 6 | 46,884 | 45,461 | 12,497 | 25 (majority) | Sep 02 - Oct 04 | 21963518 | 61379 | 1.86 | 1.79 |
| Ekurhuleni 7 | 104,407 | 100,921 | 46,503 | 25 (majority) | Sep 02 - Oct 04 | 20878052 | 60891 | 1.12 | 1.05 |
| Ekurhuleni 8 | 82,503 | 72,334 | 22,267 | 25 (majority) | Aug 02 - Sep 04 | 24136060 | 68898 | 1.18 | 1.14 |
| Ekurhuleni 9 | 64,776 | 63,812 | 49,400 | 25 majority | Sep 02 - Oct 04 | 6537073 | 19164 | 1.32 | 1.25 |

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| Johannesburg 1 | 6,594 | 6,398 | 2,923 | 25 | Feb 02 - Mar 04 | 803689 | 2332 | 0.67 | 0.64 |
| Johannesburg 2 | 41,086 | 39,699 | 23,134 | 25 | Feb 02 - Mar 04 | 8308707 | 23806 | 1.43 | 1.33 |
| Johannesburg 3 | 132,061 | 123,262 | 67,218 | 25 | Jan 02 - Feb 004 | 47344048 | 135363 | 2.32 | 2.09 |
| Johannesburg 4 | 106,225 | 96,235 | 48,743 | 25 | Feb 02 - Mar 04 | 35949232 | 103278 | 2.06 | 1.80 |
| Johannesburg 5 | 165,967 | 160,076 | 149,650 | 25 (148 057 (89% records have no readings) | Feb 02 - Mar 04 | 10238229 | 30722 | 2.29 | 1.88 |
| Johannesburg 6 | 128,483 | 118,315 | 53,668 | 25 | Feb 02 -Mar 04 | 76233480 | 287743 | 4.36 | 3.85 |
| Johannesburg 7 | 118,793 | 105,494 | 78,386 | 25 | Feb 02-Mar 04 | 21685446 | 61570 | 1.69 | 1.52 |
| Tshwane 1 | 18,063 | 17,753 | 2,899 | 36 | Feb 02 - Feb 05 | 5460691 | 15011 | 1.01 | 0.99 |
| Tshwane 2 | 37,253 | 35,652 | 1,742 | 25 | Dec 02 -Jan 05 | 26605304 | 73138 | 2.15 | 2.06 |
| Tshwane 3 | 7,557 | 5,737 | 7,552 | | | 66404 | 183 | 45.81 | 36.65 |
| Tshwane 4 | 38,642 | 38,493 | 7,184 | 36 | Feb 02 - Feb 05 | 8576809 | 23836 | 0.76 | 0.76 |
| Tshwane 5 | 88,102 | 84,900 | 19,958 | 25 | Feb 03 -Mar 05 | 18259324 | 53234 | 0.82 | 0.78 |
| Tshwane 6 | 59,171 | 45,793 | 1,263 | 27 | Dec 02 - Mar 05 | 13887147 | 38440 | 0.86 | 0.66 |
| Tshwane 7 | 174,000 | 146,691 | 43,546 | 36 | Feb 02 -Feb 05 | 98195992 | 271560 | 2.52 | 2.08 |
| Tshwane 8 | 12,231 | 12,108 | 933 | 13 | Jan 04 - Feb 05 | 3404962 | 9330 | 0.84 | 0.83 |
| Beaufort West | 7,443 | 6,941 | 878 | 13 | Dec 02 - Jan /Feb 04 | 1552374.3 | 4368.915 | 0.71 | 0.66 |
| Breede River | 20,612 | 20,079 | 9,249 | 24 | Oct 02 - Oct 04 | 2784882 | 7717 | 1.91 | 1.75 |
| Breede Valley | 15,389 | 15,075 | 2,299 | 60 (majority) range of 12- 60 | Nov 99 - Jan 04 | 8507058 | 23852 | 1.87 | 1.82 |
| Cederberg | 7,158 | 6,881 | 2,308 | 13 | Dec 02 - Dec 03 | 2253426 | 6284 | 1.37 | 8.20 |
| Drakenstein | 40,772 | 37,076 | 11,110 | 25 (majority) range of 1 - 70 | Nov/Dec 01 - Nov/Dec 03 | 13412065 | 38471 | 1.48 | 1.30 |
| Matzikama | 8,706 | 8,317 | 1,637 | 25 (majority) range of 1 - 31 | Jan 02 - Feb 04 | 4199155 | 11505 | 1.70 | 1.63 |
| Oudtshoorn | 23,366 | | 12,000 | 24 | Nov 01 - Nov 03 | 5014013 | 13803 | | |
| Randfontein | 19,304 | 17,979 | 174 | 25 | Apr 01 - May 03 | 5602136 | 15696 | 0.88 | 0.83 |
| Sedibeng | 172,383 | 170,126 | 129,554 | 25 (majority) range 1 -42 | June 02 - Jul 04 | 24441034 | 66962 | 1.64 | 1.56 |
| Stellenbosch | 15,250 | 14,335 | 3,249 | 24 | Oct 99 - Oct 01 | 6382174 | 18301 | 1.64 | 1.53 |
| Swartland | 23,652 | 22,763 | 9,018 | 24 (majority) range of 1-24 | Oct 01 - Oct 03 | 3632927 | 10080 | 0.73 | 0.89 |
| Theewaterskloof | 16,852 | 16,305 | 6,230 | 13 (majority) | Sep 03 -Oct 04 | 3138804 | 8930 | 0.88 | 1.63 |
| **TOTAL** | **2,792,053** | **2,570,396** | **1,126,530** | | | **787577186** | **2308265.9** | | |

## 2.4 Data characteristics

The data represents a good range in terms time period covered, water regions and development characteristics. Figure 1 presents the periods of available data according to economic category. It is clear from the graph that the data for most areas exceed two years.

**Figure 1: Available data periods according to economic category**




The distribution of data in DWAF water regions is shown in figure 2. The figure shows that the largest number of consumers are located in the Northern water region. This group of consumers are all in Gauteng and include the three metros of Johannesburg, Tshwane and Ekurhuleni. The next largest group is situated in the South Western region and includes the metro of Cape Town. A large number of stands (41% of the total number) were assigned as 'vacant' stands in the treasury.

**Figure 2: Distribution of normal and vacant records according to water region**

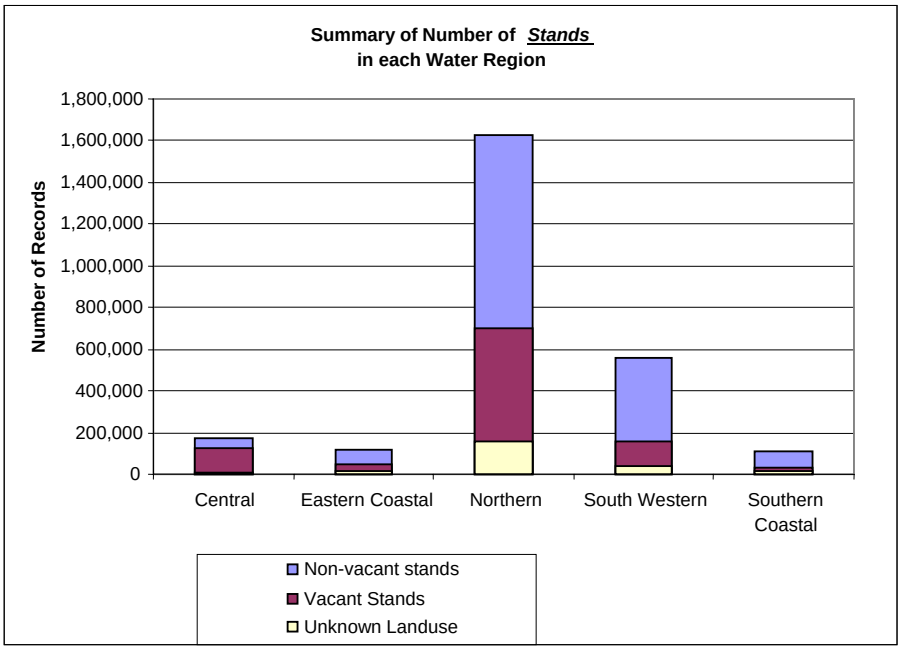
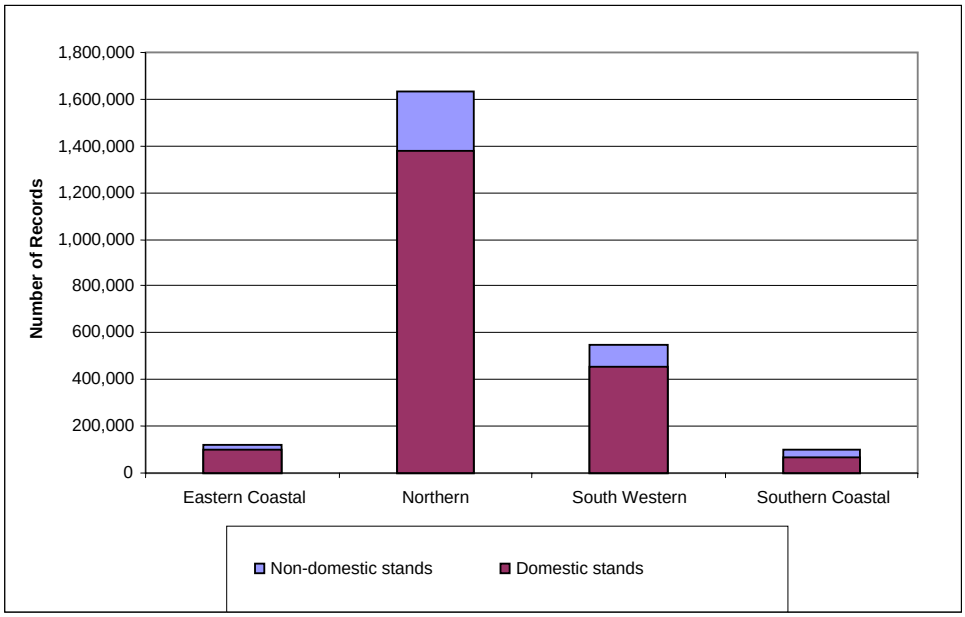



Figure 3 give the total number of stands and the number of domestic stands. It is clear that domestic stands make up the largest part of the data in all water regions. Not all of these 'vacant' stands had zero water consumption; quite a number had relevant and regular readings. It may be possible that these stands had been vacant but had since been occupied but the treasury systems had not been updated.

**Figure 3: Distribution of domestic and non-domestic stands according to water region**



The distribution of data according to size (metro or town) and location (coastal or inland) of municipality is shown in figure 4. The figure shows most data in the Inland Metro category. This is followed by Inland Town, Coastal Metro, Coastal Town, Coastal City and Coastal and Inland Town. The fraction of the data points that represent vacant stands or users are also shown.

**Figure 4: Distribution of normal and vacant records according to size and locality**

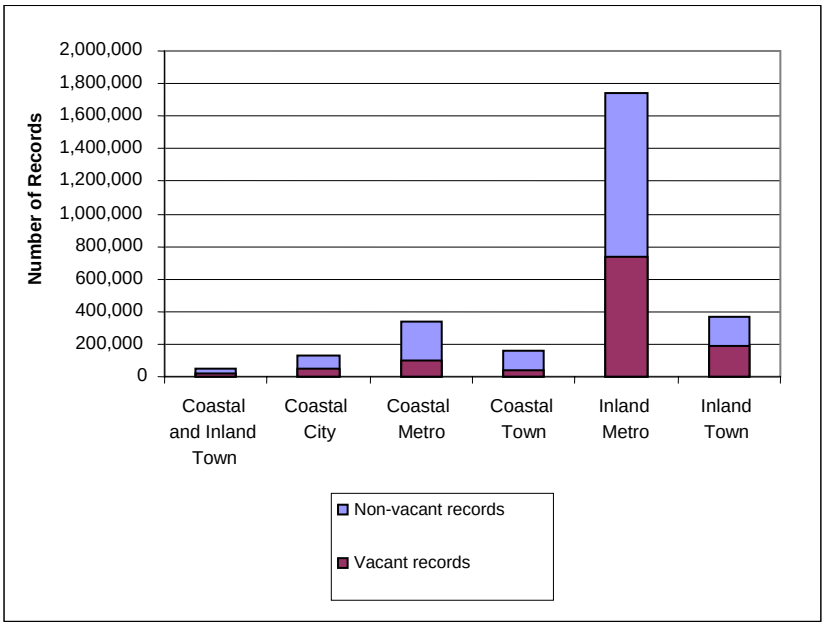
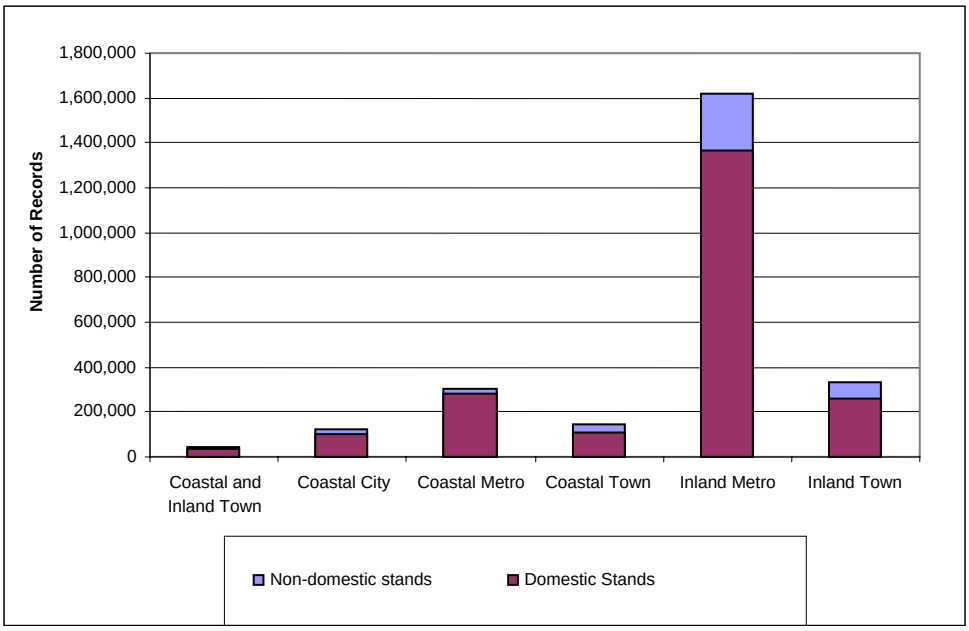



Figure 5 gives the total number of stands and the number of domestic stands. It is again clear that domestic stands make up the largest part of the data in all water regions.

**Figure 5: Distribution of domestic and non-domestic stands according to size and locality**



# CHAPTER 3: METHODOLOGY

## 3.1 INTRODUCTION

This section discusses the methodology used for abstracting and storing the data. The aim of the project is to create a data format that is software independent and easy to access.

The raw data abstracted from municipal treasury databases include many problems such as reading errors and missing data. Swift uses a number of correction procedures to correct obvious errors in the data. While the corrected data will be adequate for some types of analyses, it is not possible to predict all future applications of the data, and thus whether the current correction procedures will be adequate. For this reason the raw data is also included in the archive.

**Note: Users of the consumption archive should make themselves familiar with the data and its limitations before using it for any analyses or recommendations.**

## 3.2 DATA EXTRACTION METHODOLOGY

Municipal databases are huge and a number of well-defined procedures have to be followed to transfer the data into a Swift database and use it further. The typical steps are given below, although all of the steps are not always implemented.

1. Set up Swift data files :
    a. Copy blank data files to the correct folder;
    b. Configure the Stand ID and Land-use Categorisation codes.
2. Link the treasury data :
    a. Create a new Stands table;
    b. Set up the file references for the project;
    c. Store the treasury data files in the correct folders according to the Swift standard;
    d. Extract data from the treasury system into the Stands table.
3. Use the water meter readings to :
    a. Calculate the demand characteristics for each record;
    b. Make adjustments to the data records if required (see section 3.3).
4. Build suburb reports and calculate unit water demands :
    a. Categorise suburbs and land uses. Suburbs and land uses are mapped onto generic descriptions (e.g. suburbs X-ext 1 and X-ext 2 may be mapped to suburb X);
    b. Consolidate the number of units per stand (e.g. townhouses with many units on one stand);
    c. Identify vacant stands and large users;
    d. Generate demand reports.
5. Link the records to their GIS co-ordinates.
6. Link with Wadiso distribution system model.
7. Estimate losses :
    a. Assign each record to a bulk meter zone;
    b. Assign each bulk meter to a bulk meter zone;
    c. Compare the two to estimate losses.
8. Incorporate losses into reports and models.
9. Link with Sewsan sewer model.
10. Link demand profiles to the different user types.
11. Generate financial reports based on consumption data.
12. Advanced functions :

a. Query database;
b. Import and export data;
c. Build user-defined reports;
d. Update data.

## 3.3 DATA CORRECTION

The data used in this study was obtained from municipal meter readings. These readings are entered into treasury databases for billing purposes and were extracted using the Swift software. It is necessary to verify the data to identify and eliminate potential errors in as objective way possible. Various possible errors can be included in the database, including the following:

- Errors in dates or water meter readings
- Missing water meter readings
- Inconsistent dates
- Meter changes or clock-overs
- Meter readings not taken on the same day of the month.

Swift assigns adjustment codes to every treasury record during the process of checking the readings and calculating the annual average daily water demand and total water demand for the record. Table 5 lists the Swift adjustment codes and gives a description of the meaning of each code.

Monthly consumption figures cannot be exported directly from these corrected Swift readings due to inconsistencies in the number of days between meter readings. These inconsistencies are due to:

- The meters are not always read on the same day of the month
- In large towns meters are read throughout the month so that the period of consumption will differ for the records in the town
- In some instances there may be months when the meters are not read at all. In other cases there may be multiple readings in a month

Swift therefore interpolates readings on the 25 of each month based on the corrected readings. The monthly consumption data which are archived are calculated from the interpolated readings.

It should be noted that besides the estimated monthly consumption data, the original meter readings and reading dates are also included in the archive. Users are thus free to estimate monthly consumption figures in a different way if required.

**Table 5: Correction codes used by Swift**

| No. | Code | Description |
|---|---|---|
| 1 | None | No readings were found for the stand. |
| 2 | Single | A single reading was found for the stand. |
| 3 | Dates | Inconsistent dates. The system could not determine the order in which the readings should be placed due to date inconsistencies. |
| 4 | Clock | Clock-overs. This is where the meters have clocked over. The system will manage this situation, given the display size of the meter. |
| 5 | New Meter | This is where the new meters have been installed. The system will handle the new meter from the date it receives the first new reading. |

| | | |
|---|---|---|
| 6 | Spike | Spike interpolation. The system has detected a spike in the readings and will use interpolation to filter it out. |
| 7 | Dip | Dip interpolation. The system has detected a dip in the readings and will use interpolation to filter it out. |
| 8 | Estimate | Estimations. The system has detected that readings were estimated and will replace the estimated readings with a linear curve through the actual readings. |
| 9 | Duplicate | Duplicate dates. Two readings were found with different readings on the same date. The system will try each of the possibilities and select a suitable one. |
| 10 | Error | Reading Errors. The system was unable to determine the average water demand for the stand as the readings were totally inconsistent. |
| 11 | StartDate | Starting Dates. The starting date of the first meter reading is more than three months older than the Readings month specified. |
| 13 | RecDate | Inconsistent date recovery. The system found two readings roughly one year apart and was able to determine an AADD for the plot. The other readings were discarded. |
| 14 | User | The user adjusted the readings |
| 15 | TrimRec | Inconsistent reading recovery. The system found an inconsistent reading record, but recovered by trimming records to the first year. |
| 17 | Zero units, AADD > 0 | Unit_Ts = 0, Assumed Units = 1 for AADDPerUnit |
| 20 | Pre-paid not found | Prepaid record was not found in the stands table and hence the readings could not be updated. |
| 21 | Pre-paid code, no overlap | Prepaid record was found in the stands table but the code did not match the code specified for prepaid records. The period of pre-paid readings did not overlap with the period of the meter readings. |
| 22 | Pre-paid code, overlap | Prepaid record was found in the stands table but the code did not match the code specified for prepaid records. The period of pre-paid readings overlapped with the period of the meter readings |
| 30 | Replaced | The meter was identified as having been replaced by another meter. The 2 records were concatenated |
| 31 | Olap cons=0 | The meter was identified as having been replaced by another meter, but the record periods overlapped. The consumption during the overlapping period = 0 |
| 32 | Olap cons>0 | The meter was identified as having been replaced by another meter, but the record periods overlapped. The consumption during the overlapping period > 0 |
| 33 | Discarded | The meter was identified as having been replaced by another meter, but the record initial record was more than a user specified number of months old and was discarded. |

## 3.4 WRITING ARCHIVE FILES

The number of months of readings stored in Swift ranges from 12 to 36 months. The user has the option to archive all the data stored in Swift or only the data up until the start of the municipal financial year corresponding to the Swift project. This allows the archive to be populated with as much historical information as available in the initial archiving procedure. Updates can then be performed where only 1 year of data is stored.

Note: If the Swift project date (i.e. the month of the data extract from the treasury system) does not correspond to the end of a financial year, only the readings and consumptions that are available will be archived.

The steps in exporting an archive file are as follows:

- The Swift project should have been completed before the archiving is performed.
- Open Swift and open the project which should be archived.
- In the main menu click on Archive; Archive Data.
- A form will appear which allows the user to filter the data which is to be archived if required.
  - If no filtering is to be performed:
    - Click on the OK button.
  - If filtering is to be performed:
    - Select from the drop down list the field on which the data is to be filtered.
    - Click on the Query Button. A list of all available values for the field will be listed. Select the values which are to be included in the filtered data. Only those records with values which match the selected values will be included in the archiving process.
    - Click OK.
- Click on the folder button and browse to the folder where the archived data must be stored. The same folder must be used for all archiving of a particular town.
- Choose whether to archive only the last year of data (the default) or all the data in the folder. All the data should only be exported the first time that the town's data is exported.
- Click OK.
- If data already exists a warning will be generated saying that the data will be overwritten. Click Yes to continue if the data is to be overwritten or No to abort the archiving process.
- On conclusion of the archiving process a message will be shown stating how many records have been archived for each year of data.

## 3.5 MANIPULATING AND CHECKING ARCHIVED FILES

After writing the archive files, a number of steps have to be followed to ensure that the archive files are correct and complete:

- Check the archive data by doing spot checks on individual records (i.e. comparing Swift and archive records) and ensuring that all records defined in section 4.4 are exported.

- If a city already exists in the archive, open the last year already archived and check whether the full year's data is included. If not, these files should be replaced by the newly written archive files (assuming that the new files include the full year's data).

- Confidential and public files should be separated. Confidential files include data that is sensitive, for instance data that can be used to identify a stand location or owner. Confidential files include:
  - ID table (ID_yyyy.txt)
  - Stand link table (StandList_yyyy.txt)

  Non-confidential or public files include:
  - Consumption table (Data_yyyy.txt)
  - Reading table (Read_yyyy.txt)
  - Swift project file (SwiftProjectInfo_yyyy.txt)
  - Code table (Codes_yyyy.txt)
  - Tariff codes table (TariffCodes_yyyy.txt)
  - Rates table (Rates_yyyy.txt).

- Make backups of all new archive files.

- Provide access to the public archive files.

## 3.6 SUSTAINED ARCHIVING

The code written for exporting archive files makes this a very efficient process that can be used on future Swift databases with minimum effort. It is proposed that the archiving procedure is run at the end of each financial year and the exported files added to the database.

## 3.7 PROVIDING ACCESS TO THE ARCHIVE

The water demand archive provides a basis for future research on water demand and changes in demand patterns. The ideal would be to make the data freely available to the research community. However, this has the potential of creating a number of problems, including:

- The archive includes data of a sensitive nature, including stand owners' names and addresses. If such data is made freely available, people's right to privacy might be violated.

- The data can be analysed by a researcher to show that a particular municipality is under-performing, causing embarrassment to the municipality. The ramifications of such an event might well put the WRC and the project team in a difficult situation, since we rely on the goodwill of municipalities to continue the archiving in future.

- Water consumption data can be used by commercial research companies to estimate the income and buying power of a suburb. This information can be used for commercial benefit, for instance to determine the position of a new shopping mall, and thus the data has commercial value.

It is therefore necessary to find the right balance between the benefits of making the data available and the risk of misuse. The attorneys Smuts, Kemp & Smal (SKS) were asked to provide a practically orientated legal opinion on this matter. The remainder of this chapter is based on their advice. While the WRC is mentioned by name, the University of Johannesburg (as contractor) or GLS (as collaborating organisation) may also be applicable.

It should be stressed that the various municipalities that implemented Swift gave their permission for the data to be used for bona-fide research purposes that excludes the exploitation of the data for commercial purposes. It will thus advisable that any party requesting access to the data for the purposes of commercial exploitation first get permission from the relevant municipalities.

**Based on legal advice received, request for and permission for the use of data by other interested parties for the sole purposes of research, shall be at the sole discretion of the WRC subject to a written motivation and acceptance of an indemnity declaration (example attached as Appendix A).**

# CHAPTER 4: ARCHIVE STRUCTURE

## 4.1 INTRODUCTION

This chapter discusses the archive data structure. The structure has been designed with the following objectives in mind:

- To preserve raw data.
- To preserve adjusted or calculated data that can be of value to users, together with a description of the methods used to adjust the data.
- To provide a modular structure that allows new data to be added to the archive with least trouble.

## 4.2 ARCHIVE STANDARDS

A number of standards were set for the archive to ensure repeatability and efficient data retrieval in future:

- The unit used for building the archive is a single year's data for a given city (or town). While metros are sometimes created or sub-divided, the demarcation of individual cities and towns don't often change. The time unit of one year allows the archive to be expanded on a yearly basis by simply adding the latest year's data. It should not be necessary to change any existing files in the archive when adding new data.

- The year used was chosen to coincide with the municipal financial year, which runs from 1 July to 30 June. The reason is that changes in water tariffs and data structures are normally done at the beginning of the municipal financial years.

- Data is stored per year per town or city, e.g. Waterville/2003. The year refers to the beginning of the municipal financial year, i.e. Waterville/2003 will contain data for Waterville for the period 1 July 2003 to 30 June 2004.

- Data is stored in a non-propriety generic format as '|'-character delimited text files.

Information regarding the stand, its location, the owner and consumer is considered to be confidential or sensitive. This data is therefore stored in a file separate from the consumption and reading data, with a generated Link_ID which can be used to link the reading and consumption data back to the stand information should a suitably authorised person require this information.

## 4.3 MAINTAINING THE LINK BETWEEN DATA FILES

The intention is that new Swift data is archived on an annual basis. Readings of a particular meter are therefore split between the annual data files. In order to re-link a meter's records from the data stored in several annual files, the consumption and reading data should be archived with the Stand ID and the meter serial number fields.

As the Stand_ID field is considered confidential, a Link_ID is generated and a list of Stand_ID - Link_ID pairs is stored for the city. When the data is archived the software searches the list of Stand_ID - Link_IDs using the Stand_ID as a reference to obtain the relevant Link_ID. The Stand_ID - Link_ID link is contained in the Stands Link table. If a matching record is found the Link_ID is used as a reference in the archived files. If no match is found a new Link_ID is generated for the stand and the pair is stored in the list. Note that a stand can have more than

one flow meter, and thus a Link_ID can be linked to more than one record. The meter serial number can be used to differentiate between different meters linked to the same stand.

Both the serial number and Stand ID should be maintained since:

- The serial numbers change when a meter is replaced. In this case the link between the old and new meter will be lost, but the meter will still be associated with the same stand.

- Theoretically the meter serial number should be unique within a city. However, it has been found that in practice the serial number is sometimes not unique.

- If the structure of the Swift Stand_ID is changed (which doesn't happen often, but is possible) a new Link_ID will be created for the stand and the link will be lost. The user then has to attempt to match the records using the Serial Number field only.

## 4.4 ARCHIVE STRUCTURE

Folders and files are used to create a standard structure for the archive. Folders are used to differentiate between cities, and for each city between the years for which data is available. For example, if data is available for 2003 to 2005 for town Waterville, the folder structure will be as follows:

Archive\Waterville\2003\
Archive\Waterville\2004\
Archive\Waterville\2005\

Note that the year of the folder refers to the start of the financial year, e.g. the 2003 folder will hold data for 1 July 2003 to 30 June 2004. All the data files are placed in the lowest level folder, i.e. in the 2003, 2004 and 2005 folders for the example above. The following files are written in annual each folder (yyyy denoting the data year):

- ID table (ID_yyyy.txt): This file contains confidential or sensitive data for each record, as well as the generated Link_ID for each stand.
- Consumption table (Data_yyyy.txt): Annual consumption and other data .
- Reading table (Read_yyyy.txt): Original meter reading data for the year.
- Swift Project file (SwiftProjectInfo_yyyy.txt): A control file holding information on the archive written.
- Code table (Codes_yyyy.txt): Contains the code descriptions and mapping to categories for the various fields.
- Tariff Codes table (TariffCodes_yyyy.txt): Contains the tariff codes and mapping to tariffs.
- Rates table (Rates_yyyy.txt): Contains the rate codes, descriptions and the stepped tariff structures. The rates table is only included for years where the rates are available.
- Stand List table (StandList_yyyy.txt): Contains the mapping of the confidential Stand_ID field to a generated Link_ID field.

The files are discussed in more details in the sections below.

### 4.4.1 ID table (ID_yyyy.txt)

This file contains the confidential information for each record, pertaining to the stand, owner and consumer.

**Table 6: Fields included in the ID table (ID_yyyy.txt)**

| Column | Archive Field | Swift field | Description |
|---|---|---|---|
| 1 | ID | - | An index number for each record in the current Swift database. The index can be used to link ID, Consumption and Reading tables within a particular year, but not between different years. |
| 2 | Link_ID | - | A unique ID generated for the stand. This ID can be used to link records between years. However, since a stand can have more than one meter, various records can have the same Link_ID. |
| 3 | Stand_ID | Stand_ID | Stand ID consisting of Town, Suburb, Stand, Portion and Sub portion. The format is defined according to the treasury system |
| 4 | Unit_No | Unit_No | Additional Stand ID number to make record unique. Useful for stands with more than one meter (record). |
| 5 | Treasury_ID | Treasury_ID | Unique identification number carried by the treasury database |
| 6 | GIS_Code | GIS_Code | GIS Code (as supplied by the treasury) |
| 7 | Town | Town | The city or town. |
| 8 | Suburb | Suburb | The suburb. The suburb is mapped to a suburb category (Suburb_Category) in the Code table. |
| 9 | Stand_No | Stand_No | Stand number |
| 10 | Portion | Portion | Portion of the stand |
| 11 | Address | Address | Address of stand |
| 12 | Stand_Owner | Stand_Owner | Name of stand owner |
| 13 | Consumer | Consumer | Name of consumer (can differ from stand owner) |
| 14 | Account_No | Account_No | Municipal account number |

### 4.4.2 Consumption table (Data_yyyy.txt)

Contains the data archived from Swift as well as the adjusted monthly consumptions for the relevant financial year.

If multiple years are archived from one project the data and the number of records (except the monthly consumptions) will be identical for each of the years archived. This is because the information is a “snap shot” of the treasury system on the day that the data was extracted from the treasury database.

**Table 7: Fields included in the consumption table (Data_yyyy.txt)**

| Column | Archive Field | Swift field | Description |
|---|---|---|---|
| 1 | ID | - | An index number for each records in the current Swift database. The index can be used to link ID, Consumption and Reading tables within a particular year, but not between different years. |
| 2 | Link_ID | - | A unique ID generated for the stand. This ID can be used to link records between years. However, since a stand can have more than one meter, various records can have the same Link_ID. |
| 3 | Serial_No | Serial_No | Flow meter serial number. See section 4.3 for more information. |
| 4 | Town | Town | The city or town. |
| 5 | Suburb | Suburb | The suburb. The suburb is mapped to a suburb category (Suburb_Category) in the Code table. |
| 6 | Suburb_Category | Suburb_Category | A suburb category based on the suburb (Suburb) groupings defined in the Code table. |
| 7 | WardNr | WardNr | Political ward classification number. |

| | | | |
|---|---|---|---|
| 8 | Landuse | Landuse | Land use code assigned by the municipality. The land use code is mapped to a land use category (Land_Category) in the Code table. |
| 9 | Zoning | Zoning | Zoning code assigned by the municipality. The zoning code is mapped to a zoning category (Zone_Category) in the Code table. |
| 10 | Tax_Tariff | Tax_Tariff | Tax tariff code assigned by the municipality. |
| 11 | Water_Tariff | Water_Tariff | Water tariff code assigned by the municipality. The water tariff code is mapped to a water use category (Land_Category) in the Code table. |
| 12 | Avail_Code | Avail_Code | Water availability code assigned by the municipality. |
| 13 | Land_Category | Land_Category | A land use category based on the land use (Landuse) groupings defined in the Code table. |
| 14 | Zone_Category | Zone_Category | A zone category based on the zone (Zoning) groupings defined in the Code table. |
| 15 | Value_Site | Value_Site | Municipal valuation of the site. |
| 16 | Value_Improvement | Value_Improvement | Municipal valuation of the improvements to the site. |
| 17 | Stand_Area | Stand_Area | Area of the stand. |
| 18 | Area_Unit | Area_Unit | Units for the stand area value. |
| 19 | Route_No | Route_No | Meter reader route number |
| 20 | Measurement_Unit | Measurement_Unit | Water meter measuring unit. |
| 21 | Display_Size | Display_Size | The number of digits on the meter display |
| 22 | Meter_Installation | Meter_Installation | Date that meter was installed. |
| 23 | Units | Units | The number of living units on the stand. This value is estimated based on the area of the living units and the stand. Each stand can have more than one living units, e.g. a townhouse complex. |
| 24 | Units_T | Units_T | Number of living units connected to the record. More than one living unit can be served by the same meter. |
| 25 | Adjust | Adjust | Adjustment code indicating anomalies in the readings. The adjustment codes are explained in Table 3.1. If more than one adjustment were made to the data, only the last adjustment is shown. |
| 26 | yyyz06 | - | The calculated water consumption for each month in the municipal year, sorted from the latest to the earliest month, i.e. from June of year yyyz to July of year yyyy (with yyyz = yyyy+1). The consumption is calculated from the corrected interpolated readings stored in the Swift database. See section 3.3 for a description of these calculations |
| 27 | yyyz05 | - | |
| 28 | yyyz04 | - | |
| 29 | yyyz03 | - | |
| 30 | yyyz02 | - | |
| 31 | yyyz01 | - | |
| 32 | yyyy12 | - | |
| 33 | yyyy11 | - | |
| 34 | yyyy10 | - | |
| 35 | yyyy09 | - | |
| 36 | yyyy08 | - | |
| 37 | yyyy07 | - | |

### 4.4.3 Reading Table (Read_yyyy.txt)

Contains the readings within the relevant financial year (including the last reading taken in the previous financial year)

**Table 8: Fields included in the Reading table (Read_yyyy.txt)**

| Column | Archive Field | Swift field | Description |
|---|---|---|---|
| 1 | ID | - | An index number for each record in the current file. The index can be used to link ID, Consumption and Reading tables within a particular year, but not between different years. |
| 2 | Link_ID | - | A unique ID generated for the stand. This ID can be used to link records between years. However, since a stand can have more than one meter, various records can have the same Link_ID. |
| 3 | Serial_No | Serial_No | Flow meter serial number. See section 4.3 for more information. |
| 4 | ReadDate | ReadDate | Date meter was read. |
| 5 | Reading | Reading | Meter reading. |
| 6 | Code | Code | Municipal code for reading e.g. to indicate access problem or meter broken. |

### 4.4.4 Swift Project File (SwiftProjectInfo_yyyy.txt)

This file contains information regarding the Swift project from which the archived data was obtained.

It includes the date of the Swift project. This is the month in which the data was extracted from the treasury system, i.e. the last month for the readings contained in the Swift dataset.

The extraction period corresponds to the folder name and defines the period for which the meter readings and consumptions are included in the folder.

If the latest date of the extraction period matches the Swift Project date the rate tables are written with the annual data.

### 4.4.5 Code table (Codes_yyyy.txt)

This table contains the suburb codes in the treasury system and the mapping of each suburb code to the Suburb Category used in Swift. It also contains lists of the codes for the various fields used to determine the landuse and zoning categories and the mapping of each code to its landuse category.

**Table 9: Fields included in the Code table (Codes_yyyy.txt)**

| Column | Archive field | Description |
|---|---|---|
| 1 | Type | The field in the Swift database used for the mapping |
| 2 | Code | The code held in the treasury system |
| 3 | Description | A description of the treasury code |
| 4 | AssignedTo | The category or categories to which the field is assigned |

### 4.4.6 Tariff Codes Table (TariffCodes_yyyy.txt)

Contains the tariff code and mapping to tariffs.

**Table 10: Fields included in the Tariff Codes table (TariffCodes_yyyy.txt)**

| Column | Archive field | Swift field | Description |
|---|---|---|---|
| 1 | Suburb | Suburb_Category | A suburb category based on the suburb (Suburb) groupings defined in the Code table. |
| 2 | Code(s) | User defined code fields | The codes contained in the treasury system which define the tariff at which the consumer is charged for water consumption or water availability |
| 3 | Tariff | Tariff | The tariff category assigned to the codes |

### 4.4.7 Rates Table (Rates_yyyy.txt)

Contains the tariff codes, descriptions and the stepped tariff structures which were applied in the Swift project. Note that the Rates Table gives the rates at the time when the Swift data was last extracted. If the same Swift file is used to archive previous years, the rates will not be valid for (and the Rates Table is thus not copied to ) the previous years. As a result not all directories will include a Rates Table.

**Table 11: Fields included in the Rates table (Rates_yyyy.txt)**

| Column | Archive field | Description |
|---|---|---|
| 1 | Tariff | The Swift tariff structure |
| 2 | Upper_Consump | The upper consumption limit of the stepped tariff |
| 3 | Basic Tariff | The basic charge for water availability |
| 4 | Consump Tariff | The tariff per kl water consumed |
| 5 | Description | The description of the tariff structure |

# CHAPTER 5: CONCLUSIONS

The aim of this project was to collect existing demand data in an archive that will be accessible to researchers and other interested parties in South Africa. The objectives set for the project were achieved in the following way:

- Current water consumption data to be included in the archive was identified and collected.

- A methodology and software code were developed for abstracting and storing the water consumption data.

- The water consumption data currently available from Swift databases of various municipalities in South Africa were exported to an archive. The archive includes records of more than 2.5 million stands.

- A procedure for sustained archiving of future water consumption data was developed and new data will be added to the archive in future.

The water demand archive provides a basis for future research on water demand and changes in demand patterns. The ideal would be to make the data freely available to the research community. However, this has the potential of creating a number of problems, including:

- The archive includes data of a sensitive nature, including stand owners' names and addresses. If such data is made freely available, people's right to privacy might be violated.

- The data can be analysed by a researcher to show that a particular municipality is under-performing, causing embarrassment to the municipality. The ramifications of such an event might well put the WRC and the project team in a difficult situation, since we rely on the goodwill of municipalities to continue the archiving in future.

- Water consumption data can be used by commercial research companies to estimate the income and buying power of a suburb. This information can be used for commercial benefit, for instance to determine the position of a new shopping mall, and thus the data has commercial value.

**Based on legal advice received, request for, and permission for the use of data by other interested parties for the sole purposes of research, shall be at the sole discretion of the WRC, subject to a written motivation and acceptance of an indemnity declaration (example attached as Appendix A).**

# APPENDIX A: INDEMNITY FORM

**INDEMNITY**

*The signed indemnity must be submitted together with proof of identity, physical address (not being a post office box or poste restante) and official capacity or academic reference, as the case may be.*

I/We, the undersigned ……..
*(full names and surname)*

in my personal capacity/ in my capacity as……
*(specify position)*

of ……..……
*(specify company / entity)*

and duly authorized thereto, do hereby apply for access to the data

.....
(*furnish full details of data required, including towns/cities, years and data tables, on a separate page and annex hereto)*

in the database ……
*(describe database)*
I hereby indemnify and set free the Water Research Commission, University of Johannesburg,

GLS Consulting Engineers and ……..
*(names of clients)*
against any and all claims of whatsoever nature by any party arising from information obtained from the above database, irrespective whether such data or information in its original form or in any extracted or synthesized form prove to be correct, accurate or applicable in the circumstances and irrespective whether such data or information was published or not published.

I/We hereby confirm and warrant that:

1. The database will be used solely for the following purposes and no other, namely academic / research / design / planning other (specify other).

   ……..……

2. The database will only be used for bona-fide research purposes and neither the data nor the results of the study will be used to gain any commercial benefit.

3. Access to the database and all data therein will be restricted, securely under my control, protected (*inter alia* by passwords) and no unauthorized person or third party will be allowed any access thereto.

4. Any research, results or conclusions published on the basis of my access to the database shall be produced in a manner and format that will not allow any individual ratepayer, consumer or group of consumers to be identified, contacted or approached by any third party.

5. I accept that any data or information furnished is done in good faith, that the Water Research Commission, University of Johannesburg, GLS Consulting Engineers or the owners of the data do not guarantee or promise the accuracy or validity of the data and do not accept any liability for any information or results obtained from the data.

Thus signed at …….. on ……..200……

……
APPLICANT(S)

AS WITNESSES:

1……..

2……..