Development of a data consolidation platform for a web-based energy information system

IM Prinsloo

21660379

Dissertation submitted in fulfilment of the requirements for the degree *Magister* in Computer and Electronic Engineering at the Potchefstroom Campus of the North-West University

Supervisor: Dr R Pelzer

May 2015

NORTH-WEST UNIVERSITY
YUNIBESITI YA BOKONE-BOPHIRIMA
NOORDWES-UNIVERSITEIT

It all starts here ™

Abstract

Title:	Development of a data consolidation platform for a web-based energy information system
Author:	I.M. Prinsloo
Supervisor:	Dr. R. Pelzer
Keywords:	Data consolidation, Data warehouse, Energy information, Demand Side Management, Tax incentives, Resource reporting

Global energy constraints and economic conditions have placed large energy consumers under pressure to conserve resources. Several governments have acknowledged this and have employed policies to address energy shortages. In South Africa, the lacking electrical infrastructure caused severe electricity supply shortages during recent years. To alleviate the shortage, the government has revised numerous energy policies.

Consumers stand to gain financially if they embrace the opportunities offered by the revised policies. Energy management systems provide a framework that ensures alignment with specifications of the respective programs. Such a system requires a data consolidation platform to import and manage relevant data. A stored combination of consumption data, production data and financial data can be used to extract information for numerous reporting applications.

This study discusses the development of a data consolidation platform. The platform is used to collect and maintain energy related data. The platform is capable of consolidating a wide range of energy and production data into a single data set. The generic platform architecture offers users the ability to manage a wide range of data from several sources.

In order to generate reports, the platform was integrated with an existing software based energy management system. The integrated system provides a web-based interface that allows the generation and distribution of various reports. To do this the system accesses the consolidated data set.

The developed energy information tool is used by an ESCo to gather and consolidate data from multiple client systems into a single repository. Specific reports are generated by the integrated system and can be targeted at both consumers and governing bodies. The system complies with draft legislative guidelines and has been successfully implemented as a energy information tool in practice.

Acknowledgements

I would like to extend my gratitude to:

- TEMM International and HVAC International for funding the research and providing all the data and computational resources.
- Prof. E.H. Mathews and Prof. M. Kleingeld for providing funding, guidance and support.
- Dr. R. Pelzer for both his guidance and reviewing the document.
- Dr. J.C. Vosloo, Dr. J.N. du Plessis and Mr. P. Goosen for their insights and guidance.
- Mr. D. van Dyk for proof reading the document.
- Louisa-Mari for her love, support and endless patience.
- My parents for supporting me throughout my studies and providing me with the best opportunities in life.
- Last, but not least, to all my family and friends for their support and friendship.

Table of Contents

List of Tables

List of Figures

Nomenclature

Units of measure:

GW	Gigawatt
GWh	Gigawatt-hour
kl	kilolitre
kW	kilowatt
kWh	kilowatt-hour
m^2	square metre
m^3	cubic metre
MW	Megawatt
MWh	Megawat-hour
J	Joule
TJ	Terajoule
t	tonne
W	Watt
Wh	Watt-hour

Abbreviations and acronyms:

$ACID$	Atomicity, Consistency, Isolation and Durability
AES	Advanced Encryption Standard
$BETC$	Business Energy Tax Credit
$DBMS$	Database Management System
DoE	Department of Energy
DSM	Demand Side Management
$EEDSM$	Energy Efficiency Demand Side Management
$EMIS$	Energy Management Information System
EPI	Energy Performance Indicator
ER	Entity Relationship
ERD	Entity Relationship Diagram
$ESCO$	Energy Service Company
$ESKOM$	Electricity Supply Commission
ETL	Extraction Transformation Loading
FHS	File Hosting Service
GDP	Gross Domestic Product
$HTTPS$	Hyper-Text Transfer Protocol with SSL Encryption

IMAP	Internet Message Access Protocol
IP	Internet Protocol
IPMVP	International Performance Measurement and Verification
IST	Information Society Technologies
LEED–EB	Leadership in Energy and Environmental Design for Existing Buildings
MAP	Measurement and Acceptance Period
M&V	Measurement and Verification
NGER	National Greenhouse and Energy Reporting
NERSA	National Energy Regulator of South Africa
OLAP	On-Line Analysis Processing
OOAD	Object Orientated Analysis and Design
PDF	Portable Document Format
PGP	Pretty Good Privacy
PHP	PHP: Hyper-Text Preprocessor
PKI	Public Key Identification
POP	Post Office Protocol
REMS	Real-time Energy Management System
SABS	South African Bureau of Standards
SANEDI	South Africa National Energy Development Institute
SASOL	South Africa Synthetic Oil Liquid
SMTP	Simple Mail Transfer Protocol
SQL	Structured Query Language
SSL	Secure Socket Link
TOU	Time of Use
UML	Unified Modelling Language
URI	Uniform Resource Identifier Protocol

1

Introduction to energy data consolidation

The present energy environment in South Africa is presented in this chapter. Programmes and incentives, as proposed by government and regulating authorities, are introduced. The aim of these programmes and initiatives is to promote energy efficiency. Regulatory authorities and their various involvements in the industry are discussed. Opportunities, implications and reporting requirements that are linked to energy efficiency programmes and legislation changes are identified.

1.1 Global energy awareness

1.1.1 The focus on energy efficiency

Both established and emerging economies are influenced by delicate relationships between the supply and demand of energy resources. Globally, energy efficiency programmes are implemented to improve existing markets and to ensure sustainable development. Energy efficiency initiatives are internationally regarded as an effective strategy to promote economic growth.

The drive for increased energy supply capacity can be counteracted by reducing the demand. This offers many advantages to both energy consumers and suppliers. From a consumer point of view energy efficiency programmes fund themselves through reduced expenses associated with energy savings. The rapid implementation period of energy efficiency projects provides immediate relief to suppliers. Long-term sustainability is achievable by combining these programmes with generation capacity expansions.

Along with financial advantages, energy efficiency programmes offer sustainability and environmental advantages. Reduced greenhouse gas emissions can be achieved through energy efficiency programmes. More efficient industrial equipment, transportation services and green buildings translate to reduced energy demand, as well as reduced carbon emissions.

Governments and appointed agencies play essential roles in the energy industry. By adjusting and enforcing energy policies, developing nations can reduce energy demand escalations without impairing economic growth. Other benefits for countries include employment opportunities and less domestic dependence on foreign suppliers. An overall reduction in energy expenses and increased productivity of the economy is possible.

1.2 Energy overview of South Africa

1.2.1 South Africa's energy sources

South Africa relies on its copious mineral deposits to sustain its economy. As a result the extraction and processing of these minerals contribute significantly to the country's GDP (Gross Domestic Product). Mining and processing of minerals and their products are energy intensive by nature. According to the Department of Minerals and Energy (2005) the costs associated with mineral extraction are further elevated due to low ore concentrations and the depth of gold mines in the country.

In addition to the mining industry, the country plays host to expanding manufacturing and service industries. The manufacturing and service industries become more popular as the mining industry becomes less profitable. Manufacturing industries include coal to liquid fuel conversions, used by SASOL, and large scale smelting activities (Department of Minerals and Energy, 2005). However, the previously mentioned activities are also energy intensive and drive the need for more energy sources.

Among the natural resources, the country has large amounts of coal deposits. Coupled with the country's limited oil and natural gas reserves, coal has been used to meet the nation's energy demands. Finley, M. (2013) states that, during 2012, coal accounted for 72% of the country's energy sources, followed by 22% for oil and significantly less from other sources.

The coal industry in South Africa produces nearly 300 million tonnes of product annually. The environmental impact of such large scale operations is expected to be profound. Water pollution and air pollution are directly linked to the mining of coal. According to Lloyd (2002) domestic combustion of coal comprises as much as 30% of the particle load in South African air.

High carbon emissions are the result of heavy dependence on coal as a fuel source. South Africa has grown to be the leading carbon dioxide emitter in Africa and the 14th largest in the world (U.S. Energy Information Administration (IEA), 2014). The country agreed to the Kyoto Protocol in March 2002. South Africa was classified as a non-Annex 1 developing country and was not committed to quantified emission targets before 2012. Although no targets were set, this is a clear commitment to sustainable development and reduced harmful emissions.

1.2.2 Energy management and national authorities

South Africa's energy management plan is stated in The White paper on the energy policy of the Republic of South Africa (Department of Minerals and Energy, 1998). The policy aims to provide the nation with access to energy services while simultaneously addressing sustainability issues. Further development included The Energy Efficiency Strategy of the Republic of South Africa published by the Department of Minerals and Energy (2005).

The strategy addresses key aspects of energy structures in the country. It discusses development and efficiency targets leading up to 2015. Energy efficiency interventions are outlined in the document along with a 12% target for national energy efficiency improvement. An interim reduction target of 15% has been proposed for parasitical electrical usage (Department of Minerals and Energy, 2005).

Government plays a crucial role in ensuring the sustainable development of the nation. According to the Department of Minerals and Energy (1998) the government is required to ensure that the necessary resources are made available regarding energy management. Structures, systems and legislation have to be established and adapted to collect energy data. Development must be managed to ensure alignment with international standards and national targets.

SANEDI (South African National Energy Development Institute) was established as a state owned institute. The primary function of SANEDI is to direct, monitor and conduct energy research and to promote energy efficiency. NERSA (National Energy Regulator of South Africa) was established as a juristic person. Its mandate is to regulate the supply and prices of electricity, petroleum and piped gas.

SABS (South African Bureau of Standards), the government and the organisations mentioned previously work together to improve the country's energy environment. The role of the SABS regarding energy efficiency, is to establish standards used to govern energy efficiency development and implementation. Many standards for energy efficiency have been published. Two standards significant to this study are SANS 50010 Energy savings and SANS 50001 Energy management (Fiedler and Mircea, 2012; SABS Standards Division, 2011).

Legislation changes that promote energy efficiency have been applied to various acts. Recent adaptations include the amendment of section 12L of the Income tax act of 1962 and the National energy act of 2008. Both these changes can have a large impact on energy consumers and will be discussed in subsequent sections of this document.

1.2.3 South Africa's electricity situation

Historically, electricity prices in South Africa have been low. These low charges contributed to a competitiveness position in global markets, especially in the mining and industrial sectors where large portions of the country's electrical resources are consumed. Eskom is the largest electricity utility in South Africa. Approximately 95% of electricity consumed by the country is generated by Eskom. Due to large coal reserves most of the electricity is generated by coal power stations. This can be attributed to the relatively low coal cost as a result of the copious supply.

Eskom has a combined installed generation capacity of 44 GW. Coal fired power stations attribute to 38 GW (86%) of the capacity (Eskom Ltd., 2013). These figures seem large, but is not sufficient to supply the country's electricity needs. Two key factors attributed to the electricity shortage. Firstly, the primary producer of electricity (Eskom) failed to invest in additional generation capacity during the 1990's. Secondly, tariff increases allowed by the regulator did not account for infrastructure expansion expenses.

Although construction of two 4.8 GW power stations are under way, the lead time required to erect a generation facility is estimated at 10 years. In a growing economy the demand for electricity will also increase during the same period. The expansion constraints lead to energy supply shortages and force the utility to employ load shedding. Load shedding entails switching off large portions of the power grid on a rotary basis to manage the immediate electricity demand.

Eskom, in association with regulating authorities, has implemented a range of measures to manage electricity consumption and fund upgrade efforts. These measures include TOU (Time-Of-Use) tariff structures, energy efficiency initiatives, as well as substantial tariff increases (Hamer, W. and Vosloo, J.C. and Swanepoel, J.A., 2013). Figure 1.1 provides an estimated indication of electricity tariff increase in South Africa (Eskom Ltd., 2014; Statistics South Africa, 2005).

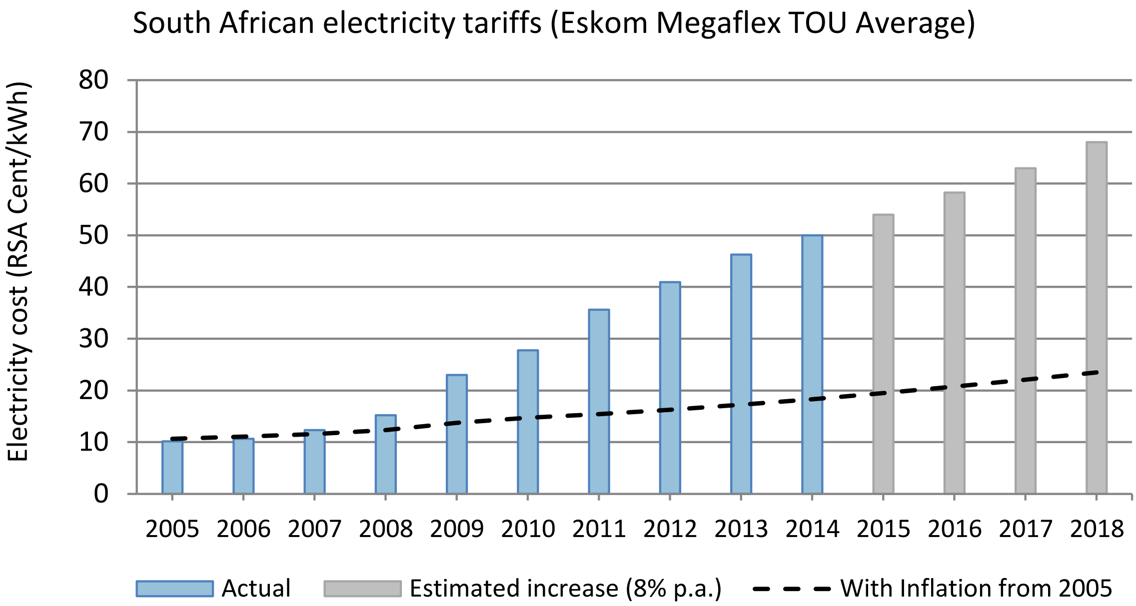


Figure 1.1: *South Africa's electricity tariffs and inflation*

From the graph it can be seen that tariff escalations surpasses inflation increases by a large margin. Profit margins of energy intensive consumers are affected by escalated bills, associated with a heavy dependence on electricity. In order to combat the negative effect of the increased expenses, consumers have the option of participating in energy efficiency programmes.

1.3 Energy saving initiatives in South Africa

1.3.1 Introduction to energy efficiency initiatives

The consequence of low electricity prices in South Africa was that there was little to no incentive to save electricity. As a result, the country has a large demand for electricity that often reaches levels dangerously close to the supply capacity (Department of Minerals and Energy, 2005). The margin between the supply and demand is referred to as a reserve margin. A reserve margin ranging from 15% to 18% is internationally accepted as standard (Department of Public Enterprises, 2007).

In 2008, the reserve margin in South Africa dropped as low as 4%. Forced load shedding had to be implemented to regulate available resources. By 2014 the country's reserve margin has not been increased sufficiently. In a June press release, Eskom Media Desk (2014) declared a state of emergency due to the severely constrained power system. In the release Eskom called upon consumers to collectively reduce system load by 10%.

Eskom projected that an additional electricity-generating capacity of 17 GW will be added to the national grid by 2019 (Eskom Ltd., 2012). Refer to Figure 1.2 (Van der Zee, 2013). From the figure it can be seen that there is a substantial electricity supply shortage. Although expansion projects are under way, intermediate relief is needed to ensure that electricity supply will not be interrupted during peak demand seasons.

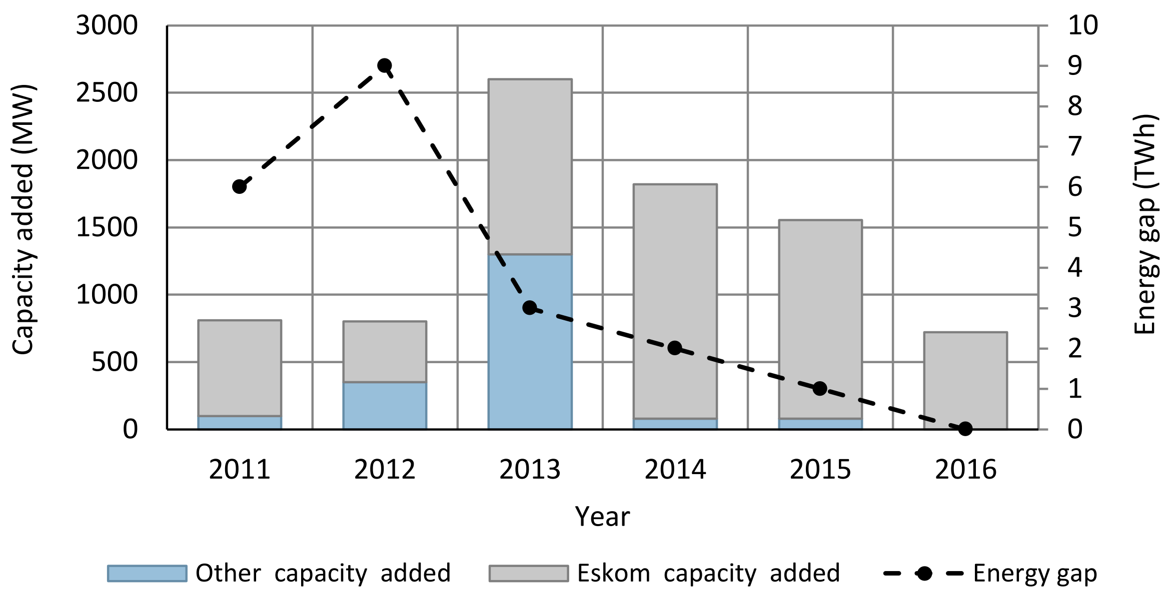


Figure 1.2: *Electricity supply expansion and shortage*

Energy efficiency programmes have grown in stature during recent years. Especially in developing countries that face the challenge of providing adequate energy services to growing populations and economies. Rapid implementation times associated with energy efficiency improvements offer attractive investment opportunities to consumers. Demand reduction projects can be implemented much quicker than supply side expansions (Spalding-Flecher, 2006).

Financial advantages serve as another motivator for energy efficiency projects. The viability of an energy efficiency project can be measured by the payback period. This can be defined as the time it takes for the avoided cost to match the initial investment cost. The projects effectively fund themselves, while offering long-term advantages to consumers and the environment.

Authorities can also utilise financial strategies to motivate energy efficiency. Penalties can be used as a short-term motivator to reduce excessively high consumption levels (Stephen, 2009). In such cases, consumers are billed according to a much higher rate . An alternative solution is rewarding consumers with rebates when they reduce their energy demand.

1.3.2 Demand Side Management interventions

DSM (Demand Side Management) implies that energy consumers undertake certain measures to reduce their load on the system. At present, Eskom and the revenue service provide funding for energy efficiency measures undertaken within South Africa. Consumers tend to show positive reaction and reduce energy consumption when financial motivations are offered.

The electricity situation in South Africa is unique considering that load reduction in the form of energy savings are considered a higher priority when compared with peak time load shifting projects. Thus, large DSM interventions are required to reduce high consumption figures and load factors. The long-term nature of the country's electricity situation is another attribute that makes it different. Intermediate load reduction can be achieved through DSM interventions, while additional generation facilities are constructed.

DSM projects can be implemented within a nine- to twelve-month period. As a result, this has the ability to provide short- to medium-term relief to the straining infrastructure. The government implemented this as a tool to grant Eskom sufficient time to complete capacity expansion projects. The Eskom DSM programme focusses on the most feasible payback period for investments (M.I. Howells, 2006).

Benefits of DSM interventions include:

- short implementation period;
- small initial investment;
- short payback period;
- positive environmental impact.

Three types of DSM project are widely implemented in South Africa. Load shifting, peak clipping and energy efficiency are the most common projects (Meek, 2013). Load shifting and peak clipping projects aim to move or reduce the amount of electricity consumed during peak periods. Load shifting projects focus exclusively on cost savings and are scaled so that the baseline and consumed energy are considered neutrally.

Peak clipping projects aim to reduce energy consumption during evening peak periods (18:00 to 20:00). Savings are calculated by comparing actual electricity consumption with a scaled baseline. The scaled baseline or service level adjustment is determined by adjusting the energy consumption baseline with a factor based on system production (Den Heijer, 2009). This is done in order to obtain baseline and actual consumption profiles that are energy neutral relative to each other.

The aim of energy efficiency projects is to reduce the overall energy consumption by enforcing a more effective control strategy. Service level adjustments are used to derive scaled baselines. Production levels and other variables such as ambient temperature are used to determine the appropriate scaling factor. Actual electricity consumption figures are then compared to the scaled baseline to determine the effective savings for the day.

The Electricity Regulation Act of 2006 entailed new regulations regarding energy efficiency and DSM (Republic of South Africa, 2006). The act was promulgated in 2008 and contained provisions for energy efficiency regulations. Aims of the policy included regulation of energy sources through governing structures, as well as the promotion of energy efficiency by means of targeted financial incentives.

1.3.3 Utilising tax to control consumer habits

Consumer habits can be controlled by adjusting prices. Worldwide taxation is utilised as a tool to guide consumers in a desired direction. Products can be made less attractive by increasing the tax to artificially increase the price. Conversely, products can be more attractive to consumers by reducing tax, thereby inducing a lower cost.

Governments also implement tax incentives to influence energy and climate change policies. Consumers are motivated to conserve energy by utilising the same tax strategies used to control product related consumer habits. The market is controlled by taxing excessive consumption of energy resources or granting tax rebates for efficiency.

The South African government applied two changes to the Income Tax Act of 1962 (Republic of South Africa, 2013b). Section 12I offers a tax rebate for a 10% energy demand reduction in the manufacturing sector. Section 12L offers further allowances regarding energy efficient savings across all sectors. Both incentives are used as tools to encourage energy efficiency and promote sustainable development.

1.4 Reporting on energy data

1.4.1 The need for energy reports

Businesses partaking in an energy efficiency initiative require consumption reports to be able to actively manage resources. Most energy initiatives rely on the same fundamental measuring system. Before any changes are made to the existing system an energy baseline is recorded. This is usually conducted by an independent third-party M&V (Measurement and Verification) specialist.

The next step is adapting or upgrading the system with the aim to improve efficiency. After implementing the upgrades, the actual energy consumption is measured and recorded. The recorded data is compared with the baseline to determine the effectiveness of the improvements. Electricity savings are typically quantified in terms of reduced kWh (kilowatt hour) usage.

The various programmes and initiatives active in South Africa have similar data requirements. The ensuing part of this section will discuss energy reporting in greater detail. Reporting requirements are classified according to initiatives and legislation policies.

1.4.2 Measurement and verification of energy reductions

Energy savings obtained as a result of the installation of DSM technologies must be verified by accredited M&V teams. These M&V teams must adhere to accreditation rules specified by governing bodies with regard to DSM interventions. According to the rules, M&V parties must be independent third-party organisations and must be accredited by NERSA.

The IPMVP (International Performance Measurement and Verification Protocol) provides an international standard for M&V procedures (Efficiency Valuation Organization, 2012). It provides an overview of techniques and practices used to verify energy saving results. The protocol does not prescribe contractual terms between parties involved in efficiency contract. Instead, it offers guidance towards the best suited measurement and verification approach for the specific application.

1.4.3 Reporting on demand side management projects

The DoE (Department of Energy) released a framework for the development of the necessary rules to give effect to the incentive scheme in May 2010 (Department of Energy, 2010). The framework discusses the SOP (Standard Offer Policy). A predetermined rate for electrical demand savings (kW) and annual electrical savings (kWh) is established through the SOP. The standard offer can be seen as a mechanism to attain demand-side resources through energy efficiency and electrical load reduction. Project developers can look forward to financial rebates based on performance.

Performance will be assessed in the form of achieved savings (kWh) for the duration of the contract. Rebates offered by the incentive will be calculated based on avoided cost of electricity supply as a result of DSM intervention. Additionally, M&V costs will be paid by the SOP administrator under the oversight of NERSA. The formula provided below shows how the rebate amount specified by the SOP is calculated (National Energy Regulator of South Africa, 2010).

$$Rebate = \frac{A}{8760 \times LF_{DSM}} + B + C - X \tag{1.1}$$

Where:

$Rebate$ = Rebate offered to consumers (R/kWh).
A = Cost of deferred capital expenditure for a coal fired power station (R/kW).
B = Avoided primary energy costs, based on electricity tariffs (R/kWh/annum).
C = Operating cost avoided (R/kWh/annum).
X = Project management and marketing cost in (R/kWh/Annum).
LF_{DSM} = Annual average load factor of EEDSM programmes (LF).

After a contract and project execution process have been established, modifications are applied to the system. Post implementation reports and supporting documentation that prove energy savings must be provided to governing bodies in order to qualify for rebates.

Project costs incurred by participating in the SOP will be reimbursed granted in the absence of verified energy savings. Energy savings are determined by comparing post implementation energy consumption of a facility with a predetermined baseline usage (National Energy Regulator of South Africa, 2010).

Every DSM project or programme must be performed in conjunction with a M&V team, in accordance with the Efficiency Valuation Organization (2012). Energy savings and peak demand savings are measured and verified to ensure that project targets have been achieved. The project developer will only receive compensation once the work has been completed and agreed energy reduction results have been verified by the M&V team.

Once the energy consumption parameters of the intervention have been established, M&V teams are responsible to audit the baseline consumption figures of the facility. After the process of installation and commissioning of approved technologies have been signed off, a MAP (measurement and acceptance period) will commence. Savings achieved during this period will be verified by the M&V entity and used as a basis for payment (National Energy Regulator of South Africa, 2010).

Recording and processing of consumption data can be seen as crucial for any DSM project. It serves not only to satisfy data needs of the M&V entity and governing authorities, but also to inform the client and project developer of performance. Performance statistics should be available on a regular basis in order to monitor and track project performance.

1.4.4 Tax incentive reporting

A tax allowance for energy efficiency savings was promulgated during November 2013. The intention is to provide encouragement to the development of energy efficiency and M&V industries in South Africa. Requirements to claim a tax allowance for energy savings in terms of section 12L of the Income Tax Act No. 58 of 1962 have been set. The act stipulates that an allowance will be granted to a tax entity, capable of proving energy efficiency savings, for a full year of assessment leading up to 1 January 2020.

As with DSM projects, an energy consumption baseline needs to be established through an accredited M&V professional in accordance with SABS Standards Division (2011). As with DSM projects, energy efficiency savings are defined as the difference between the actual amount of energy consumed during a specific period and the baseline amount that would have been used in the absence of an intervention.

Not only large scale electrical projects are eligible to qualify for the incentive. Upgrades that improve energy efficiency in buildings such as lighting, air-conditioning or insulation that result in reduced energy consumption of any source (electricity, coal, petroleum products, oil and other sources) will also qualify.

Organisations wishing to claim an energy efficiency tax allowance according to the 12L regulation need to follow certain steps. Firstly, the organisation has to register for energy efficiency tax allowance claims with SANEDI.

An accredited M&V professional needs to be appointed to compile a report. The report should contain figures relating to energy efficiency savings obtained during the year of assessment. The report needs to be submitted to SANEDI who will issue approval to commence.

The energy efficiency savings project, as well as the tax allowance approval process must be completed next. Upon completion, SANEDI will issue a formal energy savings certificate. The certificate must be submitted to the South African Revenue Service (SARS) together with a claim for tax rebate. This must be done as part of the customary tax return process (SANEDI, 2011).

The similarities between the M&V requirements for both DSM interventions and the tax rebate programme are clear. An ESCo (Energy Service Company) can therefore apply its expert knowledge to both fields. ESCos can aid with the implementation of energy efficiency projects, as well as help manage the M&V process.

The following formula will be used to calculate the deduction of taxable income for energy efficiency savings that were achieved during the year of assessment:

$$A = \frac{B \times C}{D} \tag{1.2}$$

Where:

A = Amount to be deducted (R).
B = Verified energy efficiency savings achieved during the year of assessment (kWh).
C = Applied rate as the lowest feed-in tariff expressed in (R/kWh).
D = Constant number (2) until the Minister of Trade and Industry announces change.

According to SANEDI requirements, consumed resources must be recorded and supplied to M&V personnel for verification purposes. Existing infrastructure used to report on DSM projects can be adapted to allow the import and management of tax-related data.

1.4.5 Mandatory provision of energy data reporting

One of the key challenges that the Department of Energy face, is the lack of timely and accurate energy consumption data. As a result, their annual publications are often late and feature outdated data. The lack of timely data also adds challenges to policy formulation. Timely and accurate supply, trade, stock, transformation and demand data is required when making policy decisions.

New regulations on the mandatory provision of energy data in terms of section 19 of the National Energy Act (Act No 34 of 2008) have been published. The regulations make it compulsory for consumers to report on industry specific energy consumption figures if their annual energy consumption exceeds the 180 TJ (Terajoule), or equivalent 50 GWh (Gigawatt-hour), threshold (Republic of South Africa, 2013a).

The DoE requires reports in a prescribed form and frequency. The aim of the regulation is to enable the DoE to collect, organise and publish energy information in an effective manner. This will enable the department of energy to publish accurate energy data. Furthermore, the data will be used to ensure an informed energy planning process for the country.

It is therefore imperative that these consumers record data and maintain databases that contain accurate information. To ensure that this is possible, consumers must continuously keep track of consumed energy resources. A reporting system designed with the purpose of recording and safekeeping consumption data will help to address this need.

1.5 Study motivation

A data consolidation platform is a crucial component of a system tasked with collecting or maintaining large data sets. All the reporting outcomes mentioned in the preceding sections require similar data in order to satisfy its unique requirements. The similarities make it possible to meet the discussed data requirements by using one comprehensive dataset.

This study addresses the need for a comprehensive energy data consolidation platform. Many energy management systems exist, however these systems were each developed to address a targeted purpose. These purpose orientated systems enable users to achieve specific goals by utilising included tools. Although targeted systems offer powerful solutions, it lacks the ability to adapt when system objectives change.

New legislative developments in South Africa created the need for a comprehensive energy data management platform. Electricity constraints have not been resolved in recent years and drive the need to improve energy efficiency. New energy efficiency opportunities include DSM programmes, as well as tax incentives. Alterations to the National energy act, regarding mandatory provision for energy data, have been announced by government and will be legislated in the near future.

Data consisting of a wide range of forms and frequencies need to be consolidated to comply with the outcomes of the various programme requirements and legislation. A data consolidation platform capable of accommodating this wide range of data is needed to provide timely information to M&V specialists, clients and the DoE. The comprehensive data consolidation platform developed in this study addresses the known reporting requirements.

1.6 Overview of dissertation

Chapter 1: Introduction to energy data consolidation

The present energy environment in South Africa is presented in this chapter. Programmes and incentives, as proposed by government and regulating authorities, are introduced. The aim of these programmes and initiatives is to promote energy efficiency. Regulatory authorities and their various involvements in the industry are discussed. Reporting requirements, opportunities and implications linked to energy efficiency programmes and legislation changes are identified.

Chapter 2: Data consolidation literature study

The programmes mentioned in Chapter 1 are investigated in depth. Current opportunities and technologies are investigated. Data-handling concepts and modelling techniques are considered. Data consolidation methods, as well as underlying components are identified and compared with related systems.

Chapter 3: Development of the data consolidation platform

System requirements are identified based on the findings documented in Chapter 2. A technical discussion of the software development for the data consolidation platform is provided. The development of software components are discussed based on identified system specifications. Lastly, integration with an existing reporting system is discussed.

Chapter 4: Evaluation of the data consolidation platform

The platform was developed, as discussed in the preceding chapters. After implementation the database was populated using sample data from clients in a range of industries. The developed platform was used to consolidate data into a single comprehensive data set. Results of the consolidation process are used to evaluate the efficiency of the platform.

Chapter 5: Conclusion and recommendations

This chapter provides a final discussion of the study. Completed work is summarised and discussed. Furthermore, a conclusion of the study is provided, followed by recommendations for further development in the field of data consolidation.

2

Data consolidation literature study

The programmes mentioned in Chapter 1 are investigated in depth. Current opportunities and technologies are investigated. Data-handling concepts and modelling techniques are considered. Data consolidation methods, as well as underlying components, are identified and compared with related systems.

2.1 Literature study introduction

A comprehensive literature study was compiled regarding the development of a data consolidation platform for a web-based energy information system. This chapter presents a concise account of the literature study. The research problem was divided into five secondary groups. These groups were researched in detail and form five subsections in this chapter.

Related information systems were investigated to determine positive attributes, as well as shortfalls. Data transfer channels were considered to determine the best method to transfer data between the client and the system. Data security considerations were regarded to ensure that data transfer takes place safely. Different risks and methods to avoid them were investigated.

Data consolidation considerations included the collection, preparation and storage of data. Management considerations associated with high volumes of data and heterogeneous sources were investigated. Data warehouse design and modelling form the final part of the literature study. It entails the design of system that accommodates all the identified needs and the modelling of such a system.

2.2 Related information systems

2.2.1 Energy management information systems

The International Standard Organisation announced the ISO 50001 energy management standard in 2011 (International Organization for Standardization, 2011). The standard introduced the Plan-Do-Check-Act cycle as shown in figure 2.1. This cycle ensures that energy management systems are improved on a continuous basis. Improvements can be made by reviewing reports and applying corrective actions.

The energy information system developed in this study forms a technical component of the complete Energy Management System. It can be categorised into the Check stage of the aforementioned cycle. The information system does not include features used to manage energy consumption, but rather focusses on data handling and reporting of energy performance indicators.

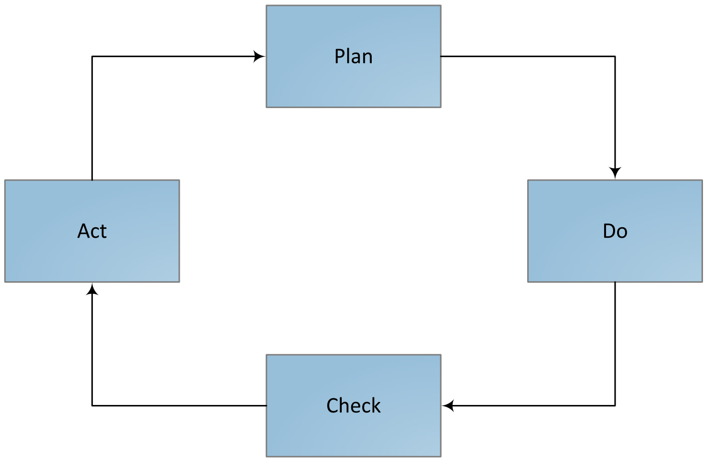


Figure 2.1: *ISO 50001 standard continuous improvement cycle*

According to the Department of Energy (2010) an energy efficiency potential of 20% – 30% exists across the country's energy portfolio. EEDSM projects still have a major role to play in reducing the electricity demand of the nation. Increasing tariffs provide ample motivation to reduce consumption. DSM programmes offer benefits to both the consumer and supplier by offering bill reductions and reduced system load respectively.

Many reporting solutions that satisfy DSM requirements are commercially available to address the need. Two similar solutions have been identified and will be presented in this section. The respective energy reporting solutions are Energy Management Information System (EMIS) from Schneider Electric (Gillis, 2009), as well as a system developed to report on mine compressed air systems by Goosen (2013).

There are many existing energy management systems to choose from. However, all systems are designed with specific purposes and have limited functionality. All the new opportunities mentioned in Chapter 1 present unique requirements. The developed data consolidation platform aims to address the identified requirements and to provide a generic platform capable of expansion.

Energy reporting systems comprise of similar fundamental components and offer comparable functions. Related systems were analysed to identify possible solutions. The best solution for the system was obtained by combining components of similar solutions. Considerations include solution functions, as well as practical implications in the South African market.

Additionally, the platform will form a single repository capable of providing structured data to a range of reports. Current applications in the South African setting include DSM performance reporting, tax incentive reporting and mandatory provision of energy data. Other solutions used to satisfy similar reporting requirement will be analysed and compared in the following sections.

EMIS by Schneider Electric

EMIS aims to promote behavioural change through awareness and accountability, identify saving opportunities, and audit savings to ensure long term persistence. Schneider's operational and maintenance support, as well as energy conservation measure identification and prioritisation programmes are used to achieve these goals.

Reports are compiled to address the following key values of interest: energy per product, energy per unit, energy usage, energy demand, financial expenditure, and greenhouse gas emissions. Different levels of detail are provided based on organisational levels. Lastly, users can define custom reports featuring the information they desire (Gillis, 2009).

This section presents three case studies for the implementation of the EMIS system. Earth Rangers in Canada, Fastweb in Italy and Sonofi Pasteur in France, are global examples of institutions that implemented the EMIS system. Their requirements and the solutions provided will be discussed. The aim of the study is to show the variety of consumer requirements across global boundaries.

Earth Rangers is a non-profit organisation involved with protecting wildlife. They identified several challenges with regard to energy management. Large consumption points in their facility had to be monitored manually. Data had to be entered and processed manually using a Microsoft Excel® Facility scorecard. Lastly, no facility to verify utility billing data existed. The manual processes were lacking and error prone, constituting the need for an automated system (Schneider Electric – Earth Rangers, 2011).

The implementation of Schneider Electric's EMIS was completed in 2011 and offered a range of solutions. Energy data from additional existing systems were imported into a central platform. A web-based interface that facilitated remote viewing of facility performance was launched. The system also allowed the client to target multiple credits under LEED-EB (Leadership in Energy and Environmental Design for Existing Buildings).

Fastweb is the second largest communications provider in Italy. They have a large optical fibre network and offer services including voice, data and video transmission. The Italian energy market underwent deregulation between 1999 and 2010. This deregulation resulted in a fragmented array of energy suppliers each with unique tariffs. Energy management software was required to meet the challenges of increased energy demand, increased tariffs and multiple supply points (Schneider Electric – Fastweb, 2010).

An extensive requirement list needed to be addressed. Energy consumption of specific entities such as buildings had to be recorded and verified against energy bills. Energy savings and set points had to be monitored. A remote system was required to manage data and support system maintenance.

Implementation of the EMIS was completed in 2010. The EMIS system was integrated with the Fastweb Ethernet Network. Monitoring capabilities for non electrical data, such as temperature and gas data, were established. A web interface provides real-time data access, which allows easy monitoring and enhanced maintenance capabilities.

Sanofi Pasteur is the vaccine division of Sanofi. Their research and development, and production facility at Marcy l'Etoile is the largest in the world and consists of more than 90 buildings. Energy data was recorded using a large network of metering devices. The network consisted of 80 electricity meters, 13 gas meters and 170 water meters. In order to preserve resources, an automated system was required to collect and process the energy data (Schneider Electric – Sanofi Pasteur, 2012).

EMIS implementation was completed in four months due to the existence of a comprehensive infrastructure. The data consolidation platform allows data collection from a range of installed devices. The data is presented in a user-friendly format, rather than raw numbers. Monthly reports were replaced with more regular reports. A securely hosted platform enables users to access data from internet devices.

DSM Monitoring system by Goosen

Goosen (2013) developed a system specifically to monitor DSM projects on mine compressed air systems. Data from mines are retrieved and processed on a daily basis. The system generates daily reports and sends it to relevant ESCo personnel and clients via e-mail. Other capabilities of the system include the generation of monthly reports and group reports.

Although the solution was targeted exclusively at mine compressed air systems, the infrastructure created the possibility to import data from a range of projects. This enabled users to track the performance of numerous DSM projects.

The system proved to be a valuable tool when used to view DSM project performance, but it had clear limitations. Data is received and processed once daily. The database accommodated only electricity data and did not make provision for consumption figures of other energy sources.

Summary of energy management information systems

The EMIS offers users a range of tools to increase productivity. System strengths include integration with existing infrastructure and user defined reporting. Furthermore, by hosting data on the internet, users are able to track consumption and performance figures remotely at a time convenient to them.

Goosen's system adequately addressed the needs of the South African market at the time that it was developed. New developments surrounding energy reporting reach further than the scope of Goosen's system. Data acquisition methods and integration with business partners in the South African market have led to a solution that is tailored to local needs.

Real-time access to data enables the platforms to rapidly identify and reflect appropriate information. However, for this to work the systems have to be integrated with existing systems on site. The success of such interactive systems will be restricted by workers who fear that automated systems will replace their occupations.

2.2.2 Tax information systems

The tax incentive mentioned in section 1.3.3 offers tax rebates to energy consumers if they comply with energy efficiency requirements. The tax rebate programme is a new incentive in the South African market. No existing solutions have been found that directly address this programme. However, similar tax rebate programmes have been implemented elsewhere.

Two United States tax programmes were considered and are presented in this section. The first is the recent Energy Efficient Commercial Deduction. This programme is very similar to the tax incentive programme implemented in South Africa. In addition to this, a case study of the BETC (Business Energy Tax Credit) programme of Oregon will be considered. The study will be used to evaluate the long term feasibility of tax incentive programmes.

Energy efficient commercial deduction

The House of Representatives of the United States of America (2005) announced that energy efficient commercial buildings deduction will be enacted as part of Title xiii–Energy policy tax incentives. The act offers up to \$1.80 of rebates per square foot (0,093 m^2) of floor space inside buildings.

Bourgeois et al. (2010) offer additional information regarding tax legislation. They state that in order to qualify for these rebates the building must achieve a 50% reduction in annual energy costs. Alternatively, system efficiency can be measured against comparable buildings that comply with standards set forth by ASHRAE (2013). The claimed savings must be verified by an IRS (Internal Revenue Service) recognised individual, using an approved computer software program.

Business Energy Tax Credit

It is clear that companies stand to gain sizable financial returns by taking part in incentive programmes. The BETC act was legislated in 1979 and has been active ever since. Each year an average of 280 businesses receive tax credits. According to Gouchoe et al. (2002), the energy savings and generation that stem from this initiative equate to roughly $100 million every year. Energy conservation projects accounted for roughly two thirds of projects submitted between 1981 and 2001.

Summary of tax information systems

From these examples it can be seen that tax incentives offer many benefits to energy consumers who participate in incentive programmes. Both programmes showed positive reaction from consumers. This proves that consumers react positively towards programmes that reduce taxation. Although the certification and auditing requirements differ from the South African programme, the same principles apply. Data must be captured and presented to authorities in a similar fashion, proving that such a system can be implemented successfully.

2.2.3 Resource reporting information systems

The Australian National Greenhouse and Energy Reporting Act was implemented in 2007 and will be discussed in this section. Various similarities exist between NGER (National Greenhouse and Energy Reporting) and the mandatory provision for an energy data reporting scheme that was proposed by the South African government.

National Greenhouse and Energy Reporting

In 2007, the Australian government introduced a energy reporting requirement namely the NGER. The NGER Act forms a single national framework. Australian corporations use this framework to report on greenhouse gas emissions, energy use and energy production. Corporations must register and provide yearly reports if they exceed the predetermined energy consumption threshold or emission threshold in a financial year.

Guideline thresholds are 25 000 t of carbon dioxide emission, consumption of 25 GWh of electricity or 2 500 kl fuel. A similar threshold mechanism applies to the South African programme. Corporations must maintain data records of their greenhouse gas emissions and energy consumption in order to compile reports in the prescribed format.

CarbonMetriX^{TM} Data Management Software by CarbonetiX (2013) is an environmental data management tool. The software is specifically designed to meet requirements set by the Australian government. It offers a secure online platform that can be used to organise and generate reports for a range of data. The software allows users to generate reports from the data set on request. The reports conform to prescribed standards specified for NGER. Data types supported by CarbonMetriX^{TM} include:

- carbon emissions;
- electricity;
- gas;
- fuel;
- water;
- paper.

2.2.4 Single repository information system

The JEVis System

The JEVis system is used as a framework for an advanced database for energy related services (Palensky, 2003). It combines data from heterogeneous sources and consolidates it in a single common database. The database forms the central point of the system and offers a range of services that stem from the central database. In addition to the data consolidation function the system offers web-enabled services to interact with the platform.

The JEVis system consists of Java-enabled web servers and an Oracle 9i relational database. A combination of server-side programmes are used to create an internet portal. This combination of software is used to integrate with other technologies and collect data. These technologies include SCADA (supervisory control and data acquisition) systems and home automation systems.

Another aspect of the JEVis system is data security. The system is used to collect sensitive client data including energy data. Basic security considerations are addressed by the JEVis system and are listed on the next page:

- secure communication channels;
- secure data storage;
- data access rights;
- smart card security systems;
- secure user interfaces.

This system is very similar to the system developed in this study. However, it is not marketed in the South African industry. The fundamental purpose of both systems are to consolidate energy related data from multiple sources into a single database and integrate with other software to allow users to access processed data.

2.2.5 Summary of related information systems

In this section related information systems were investigated. Four categories were used to discuss these system. Energy information systems included EMIS by Schneider Electric and Goosen's DSM monitoring system. Two incentives in the United States were considered as examples of tax information systems. CarbonMetrix^{TM} was considered as an example of a resource reporting system. Lastly, the JEVis system was considered as an example of a single repository information system.

Application specific capabilities of these systems were identified. However, none of the systems match all the criteria of the platform developed in this study. Features from the systems were incorporated into the developed platform. Table 2.1 summarises the capabilities of the respective systems.

Table 2.1: *Summary of related systems*

	EMIS	Goosen	US tax	CarbonMetrix^{TM}	JEVis
Designed for commercial use	Yes	Yes	Yes	Yes	Yes
Designed for industries/mines	Yes	Yes	No	No	Yes
Data transfer integration	Yes	Yes	No	No	Yes
Generic data handling capabilities	No	No	No	No	Yes
External integration options	Yes	Yes	No	No	Yes
Web integration	Yes	Yes	No	No	Yes
Scheduled reporting capabilities	No	Yes	No	No	Yes
Multi-application design	No	No	No	No	Yes
DSM reporting	Yes	Yes	No	No	No
Tax reporting	No	No	Yes	No	No
Resource reporting	No	No	No	Yes	No
Suited for South African industry	Yes	Yes	No	No	No

2.3 Data transfer channels

2.3.1 Introduction to data sharing

No analysis or reporting services are possible without relevant data. Data sharing between clients and ESCos is a crucial component of the energy management business. Data sharing discourages duplication of data collection and allows clients to disclose only authorised data to external parties. By combining the energy data with account data and production data, ESCo personnel are able to assist clients with energy related decisions.

Regular, accurate data is needed in order for an ESCo to offer dynamic services to clients. However, the disclosure of confidential data, such as production data and financial records, must be handled with care. A data sharing agreement documents which data are shared and stipulates conditions of how the data can be used. The data sharing agreement is a formal contract used to protect both parties involved in the agreement.

Along with the agreement of data composition and the use thereof, other conditions regarding the transmission frequency and scope of data content must be established. Transmission method and frequency of communication are major security concerns. Three methods used to share data over the Internet are presented in this section.

2.3.2 Data transfer via e-mail

E-mails can be used to share data over the Internet. E-mail services use established protocols including IMAP (Internet Message Access Protocol), POP (Post Office Protocol) and SMTP (Simple Mail Transfer Protocol) to relay information between end users (Eyadat and Larsen, 2011). File sharing can be achieved by attaching data files to e-mail messages. E-mails are popular due to high reliability.

However, e-mail services limit the size of messages and may not be able to handle large attachments. Confidentiality is another major concern surrounding e-mail. Non-encrypted e-mail messages can be intercepted and read or altered, without the receiver noticing. Levi and Ko (2011) state that digital certification can be used to improve e-mail security by verifying the source.

2.3.3 Data transfer via online file hosting

Cloud-based file hosting allows the transfer of large files over the Internet. According to Sanjuàs-Cuxart et al. (2012) it is fast becoming the most popular method of sharing files. Various free services offer users limited storage options. Subscription services offer more storage space and less traffic restrictions. Specialised software offers integration with client servers and offers scheduled synchronisation and backup services.

Authorised persons can then access the files over the Internet from virtually anywhere. Some hosting services send URIs (Uniform Resource Identifier) to end users. The users are directed to the files by following the link provided by the URI. Although the services claim that the URIs are secure, the files are at risk. Hosting services often rely on predictable algorithms used to generate URIs.

In a study Nikiforakis et al. (2011) used *Honey files* to measure whether attackers are abusing FHSs (File Hosting Services) to access private information. *Honey files* refer to manufactured documents that presumably contain sensitive information. The files report to a server if they are accessed. In a single month the files were accessed more than 270 times from various IP (Internet Protocol) addresses.

2.3.4 Data transfer via direct connection

Peer-to-peer file sharing refers to a direct connection between computers connected to the same network (Nikiforakis et al., 2011). Specialised software enables users to connect with specific computers using the internet. The direct connection enables users to transfer large files between linked computers. Poor security is the downside of this method. Malicious software or users can easily abuse the link to obtain private information.

2.3.5 Data transfer summary

Data transfer is a crucial component of the data consolidation platform. Data should be transferred using a reliable channel. The channel must be capable of transferring data at regular intervals. Additionally, human interaction must be avoided to ensure reliable transfer of data. Integration capabilities and automatic transfer options also need to be considered when choosing a data transfer channel.

Data transfer channels create vulnerable points in data systems. File hosting services and direct access are more vulnerable to attacks than private e-mail services. Careful consideration has to be made to ensure that data reaches the destination platform without interference. If data is intercepted unwanted data may be disclosed. This is a major concern in competitive markets and will be discussed in more detail in the following section.

2.4 Data security study

2.4.1 The Internet and data risks

Since the introduction of the Internet many systems have adapted to utilise its potential. The Internet offers users a wide network of services and platforms. One of the advantages of utilising the Internet is the ease and speed of data transfer. However, service providers do not guarantee security of hosted or transferred data. Users must therefore acknowledge the risks and take precautions regarding the safeguarding of sensitive data.

Reliable control measures must be taken when sensitive data is transferred between two entities. These control measures are necessary to maintain the integrity and confidentiality of sensitive data. To improve security many protocols and software solutions have been developed. System developers and administrators must investigate the risks involved when developing a system of a particular nature. Raja (2012) conducted a study and categorised data sharing risks as follows:

- Read - gain access to unauthorised information;
- Manipulate - adapt information;
- Spoof - provide false information or services;
- Flood - overflow computer resources;
- Redirect - change the destination of information;
- Composite - combination of the mentioned risks.

Additionally, Convery (2004) summarised the outcome of the mentioned attacks in five categories:

- disclosure of information;
- corruption of information;
- denial of service;
- theft of service;
- increased access.

2.4.2 Risks associated with file hosting

File hosting takes place through a hosted service or a cloud. Users who can provide valid credentials are offered access to the data. However, many services have only limited security measures in place and present the risk of unauthorised access. According to Gruschka and Jensen (2010) and Ramgovind et al. (2010), security is the largest shortfall of current cloud services.

Lori (2009) state that the service is only as secure as its weakest link. The lack of security of a single access point may compromise the entire cloud or hosting service where the data resides. Brodkin (2008) proposes the following list of security questions that must be discussed with cloud service providers before employing their service:

- Who has privileged access to the data?
- Is the host willing to subject to external audits?
- Does the host allow any control over the location of the data?
- Is professionally tested encryption available at all stages?
- Does the host offer file recovery, and what is the duration of such an event?
- Does the vendor have the ability to investigate illegal activity?
- What happens to the data when the vendor closes down?
- Can the data be moved to a different environment?

These questions will help to identify reliable vendors. Further considerations must be investigated with regard to data access. Gruschka and Jensen (2010) discuss potential attack surfaces involved with hosting services. Consider the cloud computing scenario shown in Figure 2.2. The three participants in this scenario are: service users, service instances and cloud providers.

Every interaction in the scenario can be linked to two participants. Additionally, every attack has two potential surfaces. The identified surfaces are:

A service-to-user;
B user-to-service;
C cloud-to-service;
D service-to-cloud;
E user-to-cloud; and
F cloud-to-user.

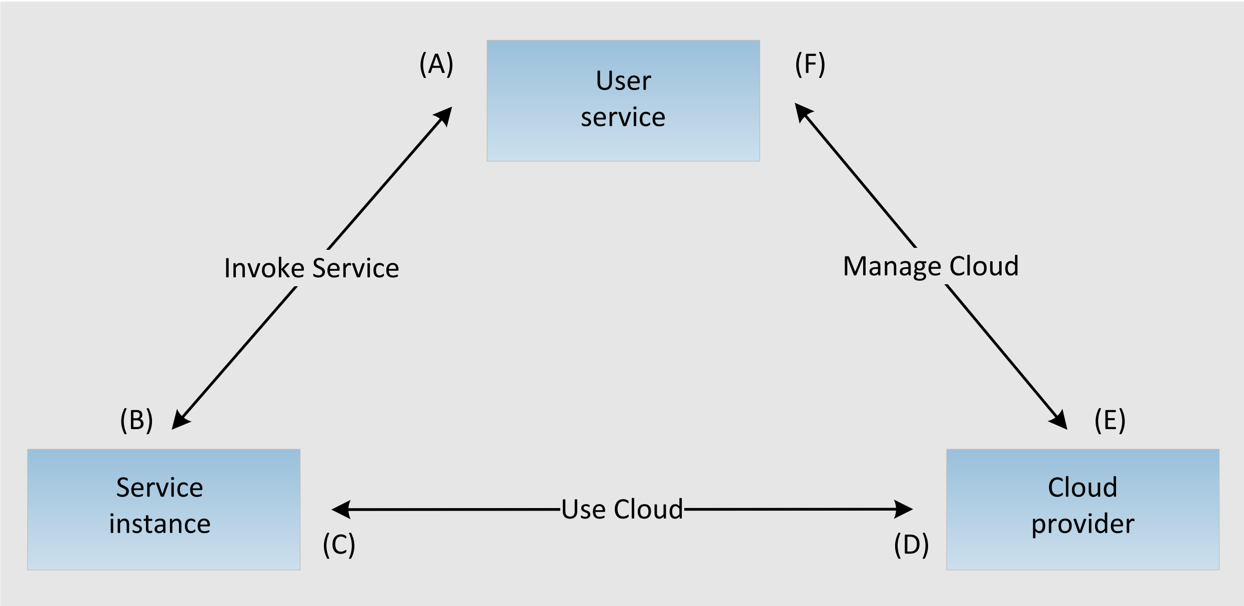


Figure 2.2: *Cloud computing attack surfaces*

Secure hosting services can improve security when using internet services. Secure services cause significant increases in system workload. Melnikov and Zeilenga (2006) support this by stating that a HTTPS (Hyper-Text Transfer Protocol with SSL Encryption) connection takes significantly more processing power and memory than standard connections. Alternatively, encryption can be used to secure data in a cloud or a hosting service. Encryption options will be discussed in more detail in 2.4.4.

2.4.3 Risks associated with e-mail messages

Typically, information is transferred specifically between a person in organisation A and a person in organisation B, or in general from organisation A to organisation B. Services such as PKI (Public Key Identification) or PGP (Pretty Good Privacy) can be employed to provide basic person to person security (Park and Devarajan, 2007). However, message transfer between organisations is much more complicated. Figure 2.3 shows an example of communication between two organisations.

Implementation of these basic authentication measures as mentioned above is not a viable security solution. This is attributed to the fact that the sender does not have direct contact with the receiver.

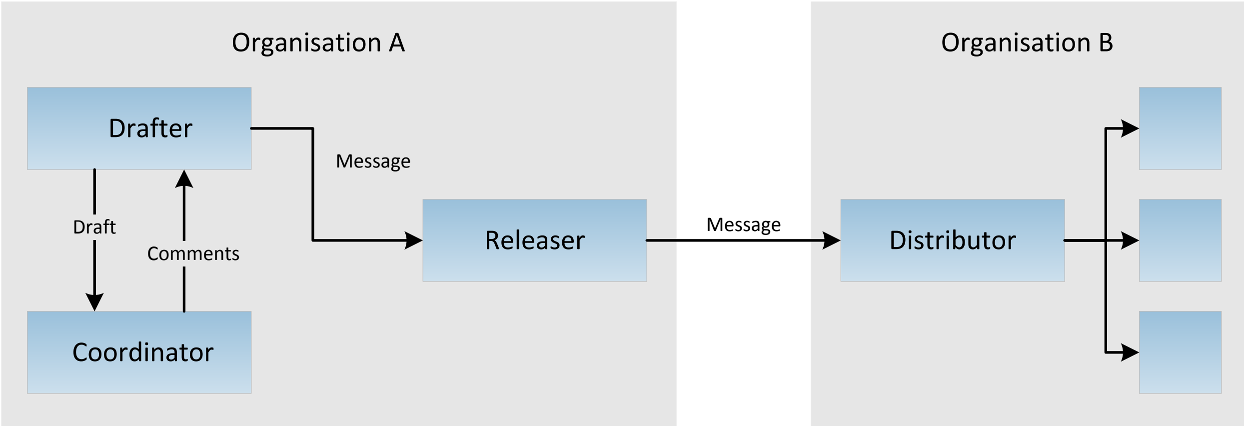


Figure 2.3: *Inter-organisation data transfer*

E-mail related risks can be divided into two categories, namely e-mail server risks and e-mail service risks. Security considerations must be taken to ensure the security of both these categories. Raja (2012) identified the following ten e-mail related risks:

- risk of network access or attack;
- mail transfer agent package risk;
- open relay risk;
- spam risk;
- virus risk;
- command abuse risk;
- mail header misuse risk;
- insecure webmail risk;
- unsafe IMAP of POP3 servers; and
- unsafe message content risk.

System developers or system administrators must evaluate the risks surrounding mail servers and messages. Various firewalls and protection software packages are designed to safeguard these systems and are available to consumers. Park and Devarajan (2007) discuss guarding software between users or the system and the internet. The guard software verifies that the content is safe to distribute or receive. Encryption is another alternative security measure that has been implemented with great success (Jain and Gosavi, 2008). The next section will look at encryption in more detail.

2.4.4 File encryption

Patel et al. (2013) defines cryptography as the science that studies mathematical techniques for keeping messages secure and free from attacks. The process of encryption and decryption can be seen in Figure 2.4. Firstly, a secret key is used in conjunction with an encryption algorithm to convert the data from plain text to cipher text. The cipher text is unintelligible and must be decrypted using the correct key and algorithm in order to reconstruct the original plain text (Asif Mushtaque et al., 2014).

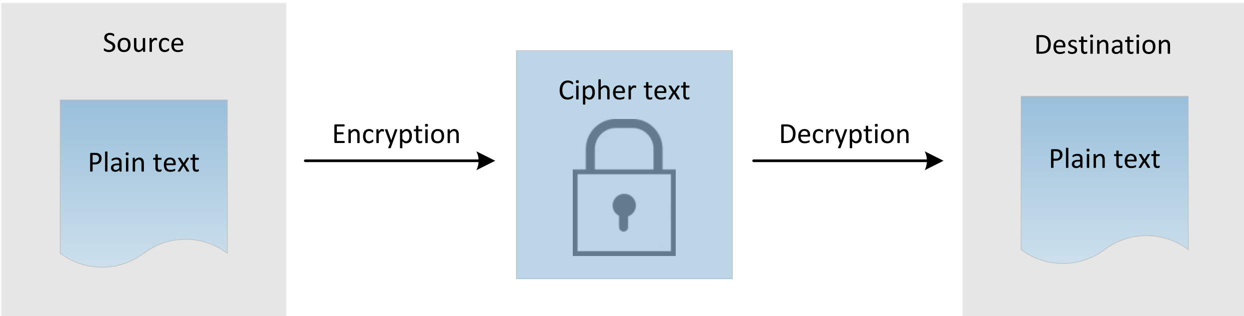


Figure 2.4: *Encryption and decryption process*

Many encryption algorithms have been developed and are available widely. These algorithms can be categorised into two primary divisions, namely symmetric-key cryptography and public-key cryptography (Chaudhari and Patel, 2014). Symmetric-key cryptography, also referred to as private key encryption, can be subdivided into block ciphers or stream ciphers. Figure 2.5 provides a visual representation of the classifications.

Jain and Gosavi (2008) propose encryption as an option to safeguard e-mail contents. Encryption can be implemented to provide similar security to files uploaded to a cloud service. Private-key and public-key encryption methods can be used to secure messages and files. However, the set-up process involved with public-key encryption, combined with the additional overhead of distributing and maintaining keys, is too complicated for the average user. Therefore, private key encryption is the preferred encryption method.

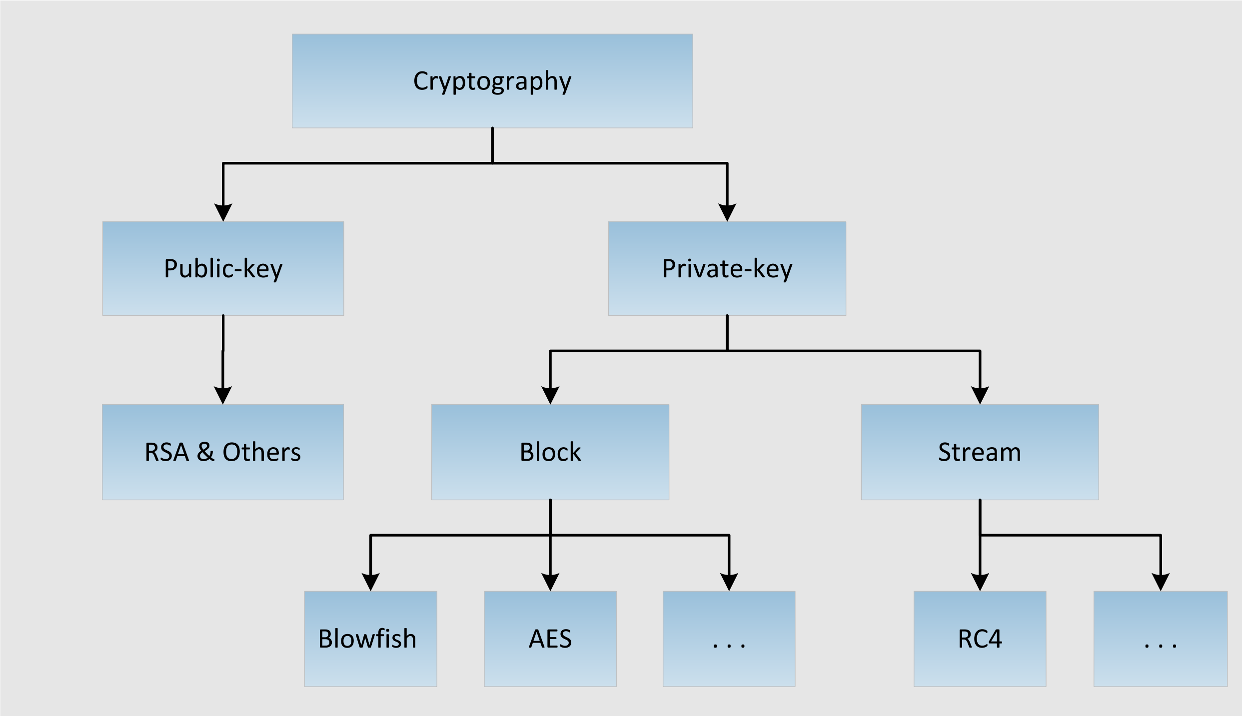


Figure 2.5: *Encryption classification*

The two most popular private key encryption algorithms will be compared throughout the remainder of this section. The two algorithms considered are AES (Advanced Encryption Standard) based on the Rijndael algorithm, as well as the Blowfish algorithm. Both these algorithms have positive and negative aspects and will be discussed in more detail.

According to Shao et al. (2010) AES is based on a submission and permutation network. The algorithm makes use of a fixed block size of 128 bits and accepts key sizes of 128, 192, or 256 bits. The algorithm is executed a fixed number of rounds depending on the key length. The algorithm will be executed 10, 12 and 14 times for keys with a length 128 bits, 192 bits and 256 bits respectively. Each round of processing involves the following steps (Chaudhari and Patel, 2014):

- divide plain text into blocks and substitute bytes within these blocks;
- shift rows according to a simple permutation;
- mix columns by multiplying the elements contained in the shift row with the algorithm's matrix;
- add the round key to the data through an exclusive or operation.

Blowfish is another symmetric block cipher. The algorithm is a Freisel Network that iterates through an encryption function 16 times. It makes use of a 64-bit block and accepts variable key lengths ranging from 32 bits to 448 bits. The algorithm can be divided into two parts, namely a key expansion part and a data encryption part (Asif Mushtaque et al., 2014).

Blowfish is often preferred over alternative algorithms such as AES because of superior results during speed tests (Elminaam et al., 2010).

According to Singh and Maini (2011), AES is flexible and can be implemented on a range of devices. The only known attacks that have succeeded against AES and Blowfish algorithms are brute force attacks. Chaudhari and Patel (2014) mention that Blowfish is considered insecure for many applications and must be improved with additional safety measures. Blowfish is considered to have superior performance when evaluating execution time.

Although Blowfish offers faster execution times, cipher text derived using this algorithm is larger than the AES alternative. This can be attributed to the difference between key size and block size (Asif Mushtaque et al., 2014). Therefore, AES can be considered superior when transmission of data is considered, because data transmission requires more resources than computation (Hirani, 2008). Table 2.2 provides a size comparison of files generated using different encryption algorithms.

Table 2.2: *Encrypted file size comparison*

Algorithm	Size before encryption	Size after encryption	Size after decryption
DES	240 KB	328 KB	240 KB
TDES	240 KB	614 KB	240 KB
AES	240 KB	847 KB	240 KB
Blowfish	240 KB	955 KB	240 KB
Twofish	240 KB	955 KB	240 KB

2.4.5 Summary of data security considerations

Security risks were investigated and categorised into three sections namely: internet risks, file hosting risks and e-mail risks. Various vulnerabilities were identified and discussed. Many risks can not be avoided, because of the system's dependence on internet services.

Reasonable data security can be achieved through careful consideration of these risks. Security measures must be implemented to ensure that limited persons have access to privileged data. Data sharing policies need to be enforced to ensure secure data transmission.

Encryption can be used when sharing data. This will render files useless to most attackers who obtain data using malicious methods. Attacks exists to decrypt encrypted files, but it takes long and may cause a large enough delay to render the data useless. AES offers strong encryption capabilities and is a widely used standard. It can be integrated with most systems by simply using existing libraries.

2.5 Data consolidation considerations

2.5.1 Introduction to data consolidation

The purpose of data consolidation is to collect, prepare and store data. Cases such as the one presented in Copin et al. (2010) involve many stakes. High volumes of capitalising data, seasonal data, as well as data streams need to be processed. The data are typically collected from various sources and are of a heterogeneous nature. The consolidation software is tasked with restructuring the data to an acceptable state for storage.

To effectively manage the vast amounts of data, EPIs (Energy Performance Indicators) should be identified. These performance indicators must be measured regularly. The measurements must then be compared with baseline measurements to evaluate system performance (Velázquez et al., 2013). Another frequently discussed concept is data mining. Hooke et al. (2004) define data mining as the process of extracting comprehensive information from immense datasets to improve business decisions.

2.5.2 The data consolidation process

Vassiliadis et al. (2002) define ETL (Extraction-Transformation-Loading) as a software set responsible for the extraction of data from multiple sources, its cleansing, customisation, reformatting, integration and insertion into a data warehouse. From the definition the process can be seen as very important, however, very little research is available in the field. Kimball and Caserta (1996) propose that this is due to the absence of a modelling standard representing the ETL activity between data sources and the data warehouse.

Both Vassiliadis et al. (2005) and Simitsis et al. (2005) describe the functionality of ETL tools by discussing its five most prominent tasks. It is responsible for:

- identifying of relevant information at the source side;
- extracting process involved with gathering the identified data;
- integrating the collected data and restructuring it to a common format;
- cleansing the data to conform with business rules;
- finally, loading the data into the database or data warehouse.

A supportive description of the ETL process in three steps is provided by El-Sappagh et al. (2011). The proposed steps are extraction, transformation and loading respectively. During the extraction phase, data presented in various formats are collected from a number of sources. The transformation step cleanses the data to a correct, complete, consistent and unambiguous format. Lastly, the data is loaded into dimensional structures and stored. The visual representation of the process can be seen in Figure 2.6.

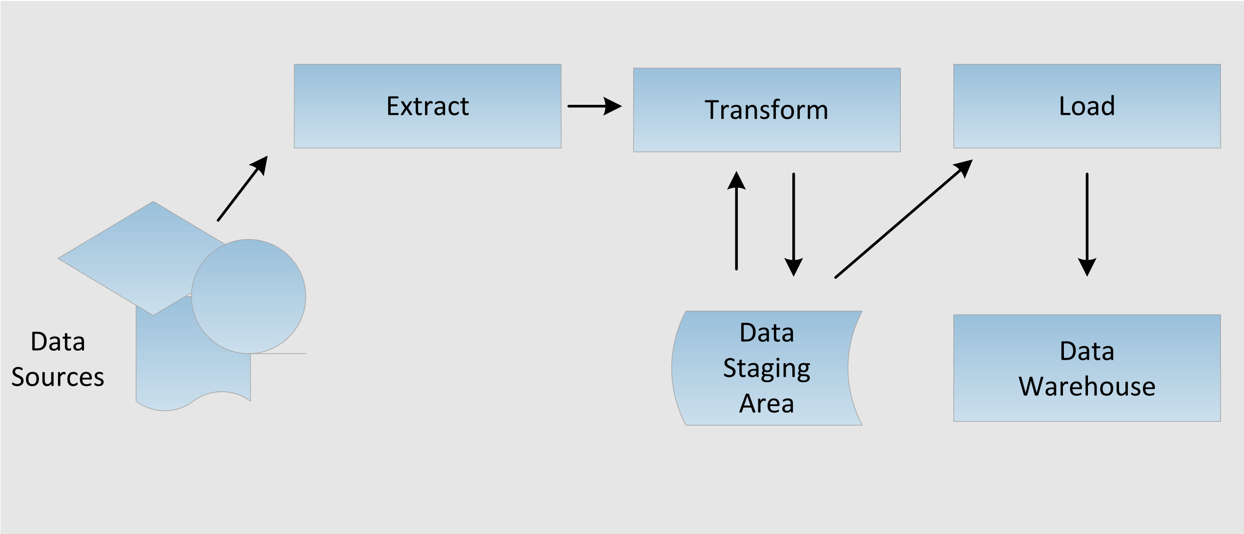


Figure 2.6: *Extract, transform and load process*

Six requirements for an ETL system can be derived from the model, proposed by El-Sappagh et al. (2011), namely:

- support integration from multiple sources;
- is robust with regard to changing sources;
- provides support for flexible transformations;
- can be employed in a suitable environment;
- is capable of handling various ETL operations;
- is simple in creating and maintaining.

The ETL process is usually executed automatically by the system and not by user guided sessions Vassiliadis et al. (2005). Typical systems will employ batch processing techniques to run the process on a scheduled basis. This enables system administrators to create backup files while the system is in a consistent state.

2.6 Data warehouse design and modelling

2.6.1 Data warehouse fundamentals

Felden and Chamoni (2003) define a data warehouse as a subject orientated, integrated, non-volatile, and time variant collection of data. A data warehouse consists of architectures, algorithms and tools used to consolidate data from multiple sources. The consolidated data is stored in a single repository that can be queried and analysed (Cui and Widom, 2003).

The power of a data warehouse lies in the fact that heterogeneous systems are combined to form a single comprehensive repository. The warehouse is used to effectively collect and manage large sets of data. Advanced data analysis is possible, when combining the warehouse with an OLAP (On-Line Analysis Processing) tool (Gökçe et al., 2009).

The design of warehouse systems is complicated due to the size and complexity of the systems with which it must be integrated. The systems are often required to be scalable and must provide consistent and reliable operations from a heterogeneous range of data. Hoang et al. (2012) acknowledges that the design and maintenance of such a complex system can be challenging.

The following list of steps is provided to aid with implementation of a data warehouse:

- analyse the critical subject area;
- study existing models and data needs;
- design a logical data model and physical database;
- design back end services and applications;
- deploy database implementation and initiate the system with starting values.

2.6.2 Database design of a data warehouse

From the preceding section it is clear that many elements have to be considered when designing a data warehouse. The most important element is the database itself. Database design influences both data import and data extraction tasks. According to Inmon (2005) the quality, correctness and completeness depend on how well the data model is structured.

Inmon (2005) conducted an exhaustive study of available database design patterns. In this section the general star design pattern will be considered. According to Blaha (2010) data warehouse applications make use of the star schema pattern frequently. The pattern is structure orientated and represents data as facts associated with dimensions. Facts are used to measure performance and dimensions are used to describe the facts.

The structure is simple to implement and allows easy access to ad-hoc queries. Figure 2.7 shows a visual representation of a star schema (Inmon, 2005). The primary keys of the dimension tables are used to form a composite primary key in the fact table. Fact data is entered at that location and is specified by clearly defined dimensions. This design allows users to select specific fact data based on a combination of dimensions.

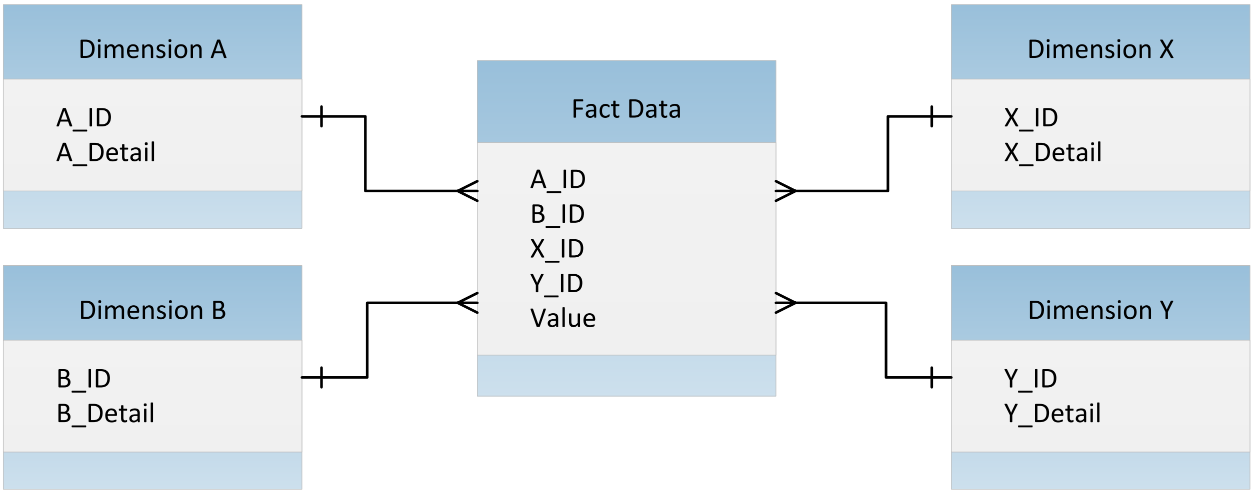


Figure 2.7: *Star database schema*

To improve data recovery and make multi-user access possible DBMS (Database Management System) employ a set of rules referred to as ACID (Atomicity, Consistency, Isolation and Durability) (Frank, 2011). These rules ensure that the database remains in a consistent state. The four elements along with referential integrity will form the discussion throughout the remainder of this section.

Atomicity guarantees that either all of the updates of a transaction are committed or that nothing is committed. This property makes it possible to re-execute failed transactions. The consistency element ensures that the database is in a valid state after completion of a transaction.

Isolation ensures that execution of concurrent transactions result in a similar manner as sequential transactions. The isolation element is important in situations where more than one user has access to the system at a given time. Durability ensures that committed transactions are stored. The history of these transactions will remain unchanged in the event of system failure.

Referential integrity is used to link figurative parent and children and ensures that no child can exist without a parent. Two options are available when enforcing referential integrity. The first option removes all referenced child entries when parent entries are removed. The other option prevents the removal of parent entries until all the referenced child entries are removed. If the database is stored in a single location this local integrity control is sufficient (Frank, 2011).

2.6.3 Modelling the database

The original ER (Entity-Relationship) model was put forward by Chen (1976). Since then many contributions have been appended to add further semantics to the model. UML (Unified Modelling Language) was developed as a standard modelling language to support OOAD (Object Orientated Analysis and Design) (Fowler, 2004). In recent years UML has evolved and is used as a popular notation when designing conceptual databases.

Figure 2.8 shows an example of a UML class (a) next to an ERD (Entity Relationship Diagram) entity (b). Attributes listed in both diagrams are listed in plain text and define properties of the entity.

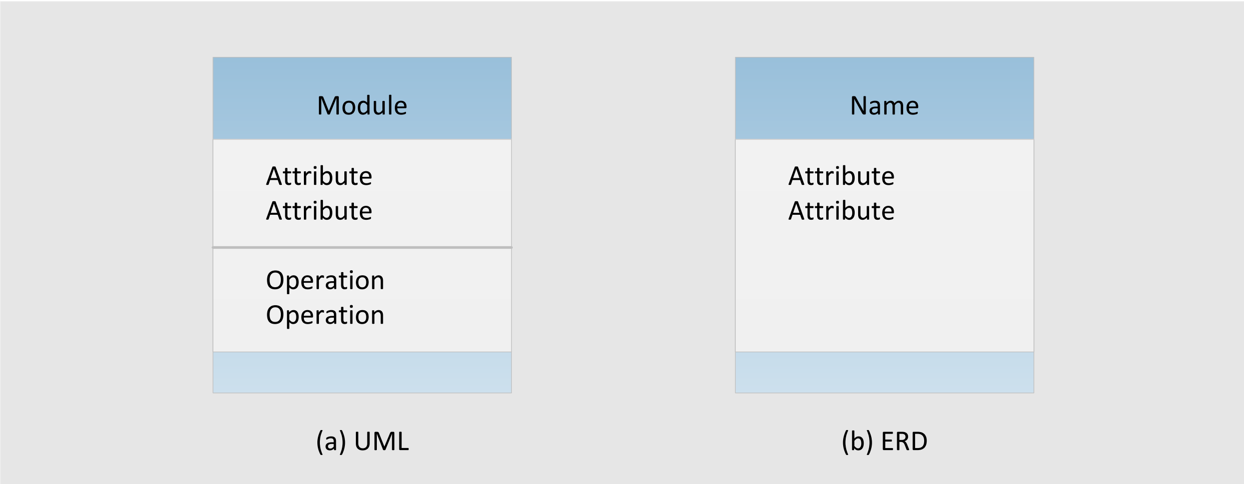


Figure 2.8: *Comparison between UML and ERD*

Properties contained with entities are linked to one another with lines. Figure 2.9 provides a visual representation of linked entities. The link shown in (a) is compulsory and is indicated with a solid line. The link seen in (b) is optional and is indicated by a dashed line.

Further detail surrounding the relationship between entities are provided by using different line ends. Refer to Figure 2.10. In this figure (a) represents a one-to-one relationship between entities, while (b) shows a one-to-many relationship.

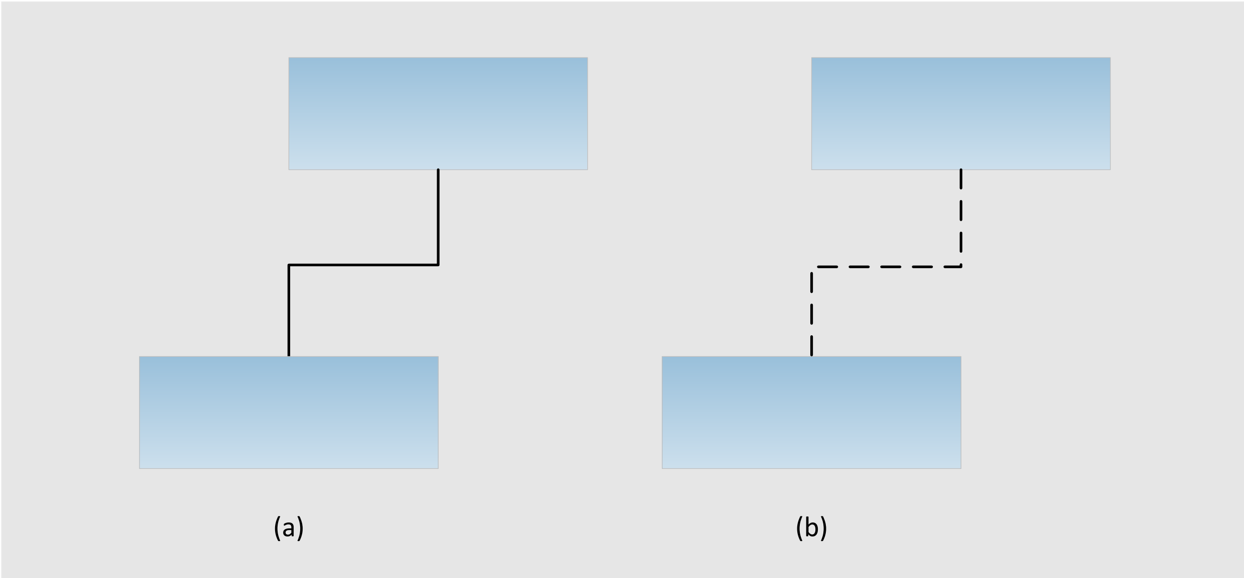


Figure 2.9: *Comparison between compulsory and optional lines*

Aggregation and composition are used to further describe a relationship between entities. Both these relationships link an auxiliary entity to a whole entity. If the whole entity is destroyed and the auxiliary entity can still exist and function on its own, the relationship is called aggregation. In the alternative case, the auxiliary entity forms part of the whole and ceases to exist if the whole entity is destroyed. Such a relationship is referred to as composition.

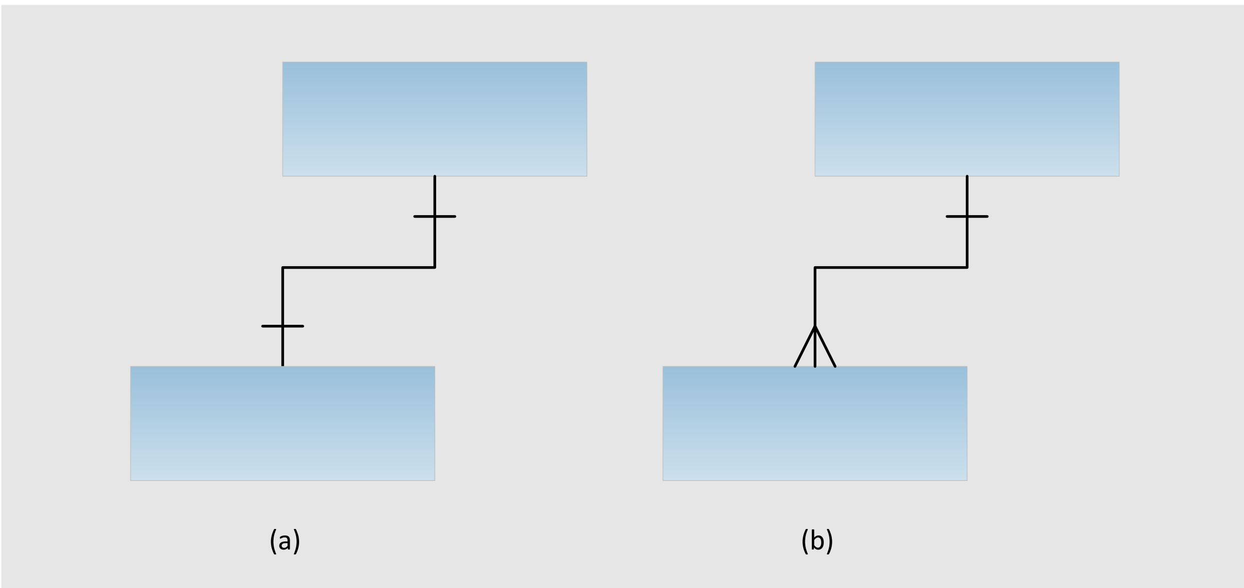


Figure 2.10: *Comparison between one-to-one and one-to-many links*

2.7 Summary of literature study

Related information systems were considered based on four categories. The categories were energy information systems, tax information systems, resource reporting systems and single repository information systems. The systems were compared based on their respective capabilities. The combined set of systems provided a good indication of required system capabilities, however none of the systems addressed all the system requirements.

Data transfer focussed on finding a reliable channel to transmit data from the client to the developed system. The aim was to identify a channel that does not require human interaction. Email was chosen as the preferred method because of its stability an security advantages over the other options that were regarded.

The developed system handles confidential client data and must therefore have data protection elements in place. A range of risks and methods to avoid them were discussed. Data encryption and security procedures provide sufficient levels of data security.

Data management systems and the modelling of such systems were investigated. The research was used to conceptualise a solution based on similar structures. Modelling tools were identified and later used to create the developed solution.

3

Development of the data consolidation platform

System requirements are identified based on the findings documented in Chapter 2. A technical discussion of the software development for the data consolidation platform is provided. The development of software components are discussed based on identified system specifications. Lastly, integration with an existing reporting system is discussed.

3.1 Introduction to platform development

The previous chapters provided background information for the study. Chapter 1 focussed on the problems at hand, while Chapter 2 considered similar problems and possible resolution methods. In this chapter the development of the new system will be discussed. The contents of the discussion include system specifications, the design details, as well as a discussion of the implementation process.

A rapid prototype life cycle methodology was used to develop the platform. Development started with an examination of related systems, followed by the development of the initial prototype database. Next, supportive software structures were developed and integrated with the existing energy management system. An iterative refinement process was used to address changing requirements and preferences.

The initial structure was adapted to deliver a generic framework, which offers extensive data handling capabilities. The generic database is coupled with a software structure that can adapt by adding modular software components. This enables the system to adapt and expand when requirements change.

Specialised process modules are integrated with the system framework as shown in figure 3.1. The generic framework is indicated in blue, while specialised models are shown in grey. These specialised models have the ability to adapt to allow unique data handling capabilities while not influencing the primary structure.

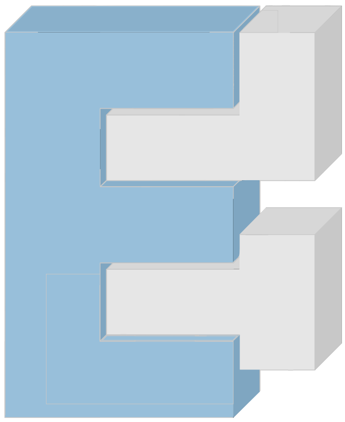


Figure 3.1: *Modular software framework*

The developed platform will be discussed in three parts. The database structure will form the first part of the discussion. Supportive software modules will form the second part of the discussion. Lastly, the discussion will include a section dedicated to the implementation process and integration with the existing system.

System specifications were established before development was possible. A number of system constraints exist due to integration with the existing energy management system. These constraints force certain design decisions into a specific direction. System specifications were derived by considering these restrictions, in combination with the contents of Chapter 1 and Chapter 2. The following section will provide a comprehensive list of specifications.

3.2 System specifications

3.2.1 Identified constraints

The developed data consolidation platform forms the backbone of the energy information system. An existing web-based energy management system is used as a starting point. The developed system must therefore integrate with the existing system. Integrating the solution with the existing system eliminates a large volume of redundant work and provides users with a familiar interface.

However, integration with an existing solution results in compatibility constraints. In order for the developed platform to integrate with the current system two critical system constraints were identified. First, the DBMS is limited to MySQL. The second constraint limits the scripting language to PHP (PHP: Hypertext Preprocessor).

While developing the existing system Goosen (2013) identified MySQL as a suitable DBMS solution. MySQL was chosen as DBMS for this solution as well, due to sufficient functionality and integration prospects. Table 3.1 shows a comparison of available database management systems.

Communication channels with clients are limited, due to security policies and intellectual property concerns. The existing system used e-mails to transfer data between clients and the ESCo. To avoid unnecessary conflicts and human relation matters, these established communication channels were chosen to transfer data to the new system as well.

Table 3.1: *Comparison of database management systems*

Database	Edition	Cost	Licence	Database size	Security	Support
MySQL [1]	Community edition	Free	GPL	Unlimited	Medium	No
	Standard edition	US$ 2 000 p.a.	Proprietary	Unlimited	Medium	Yes
PostgreSQL [2]	-	Free	PostgreSQL	Unlimited	High	Optional
Oracle [3]	Express edition	Free	Proprietary	4 GB	Medium	No
	Standard edition	US$ 17 500 p.a.	Proprietary	Unlimited	High	Yes
SQLite [4]	-	Free	Public domain	128 TB	Low	Optional
Microsoft SQL server [5]	Express edition	Free	Proprietary	4 GB	Medium	No
	Standard edition	US$ 7 172 p.a.	Proprietary	524 PB	High	Yes

The developed platform must address known system requirements. Identified requirements can be placed in four categories, namely: general system requirements, Demand Side Management requirements, tax incentive requirements and lastly, resource reporting requirements. The list in section 3.2.2 provide more details for each of the mentioned categories.

3.2.2 Platform requirements

A list of platform specifications is provided below:

- General platform requirements:
  - The platform should accommodate energy data.
  - The system must be capable of storing production data.
  - Data sampled at fixed intervals must be accommodated.
  - Budget and actual data must be stored for comparison.
  - Original files must be archived for M&V purposes.
  - The platform must accommodate a range of data sources.
  - The platform must be capable of accommodating new processing extensions without modifications to its framework.
- DSM requirements:
  - The platform must be able to store electricity data.

[1] MySQL (2014)
[2] PostgreSQL (2014)
[3] Oracle (2014)
[4] SQLite (2014)
[5] Microsoft (2014)

  - The platform must accept electricity data in the form of energy data and power data.
  - Data samples for numerous intervals during a single day should be accommodated.
  - Budget and tariff data must accompany measured data and should be handled.
  - Relevant production and climate data should be stored for support purposes.
- Tax incentive data requirements:
  - The platform must be able to store energy data.
  - Energy data is not limited exclusively to electricity but all energy data types.
  - Energy baselines, budget data and actual measurements must be stored.
  - Data is not limited exclusively to electricity but all energy data types.
  - Various data intervals should be accommodated, such as monthly and daily figures.
- Resource reporting requirements:
  - Actual energy data and production data must be handled.
  - Data is not limited exclusively to electricity but all energy data types.
  - Various data intervals should be accommodated, such as monthly and daily figures.
  - The system should be able to handle various measurement units.
  - The platform must be capable of consolidating the measurements to a single quantifiable total.

3.3 Database design

3.3.1 Development of the generic database

A new database structure had to be developed to allow generic data handling. The developed generic structure allows users to import and store a wide range of energy and production data. The structure was modelled by creating multiple tables and links to form a relational database model. Initial values were entered into the tables. Unique system modes are obtained by selecting different combinations of these values.

MySQL Workbench was used to develop ERD diagrams of the new database. The software offers powerful tools that can be used to add components included in a relational database model. The diagrams were then synchronised with the database. The created tables form the generic database structure of the data consolidation platform. The remainder of this section will stage the developed structure.

3.3.2 The generic tag

The key to the generic data consolidation platform is to break imported data down to components that are available on a universal level. Imported data is studied to find identifiers and corresponding values. Tags are used to specify attributes associated with the identifiers. These tags form the core elements of the data structure and are used to identify data elements throughout the platform.

Figure 3.2 shows the structure of the *GenData_Tag* in combination with associated property tables. Regard the *GenData_Tag* table, the first four fields are used to store unique descriptive elements of the tag. The *t_ID* field is used to identify the tag and link it to various other tables in the database. The last four entries in the *GenData_Tag* table are foreign key entries. These fields link the tag to records contained within the other displayed tables.

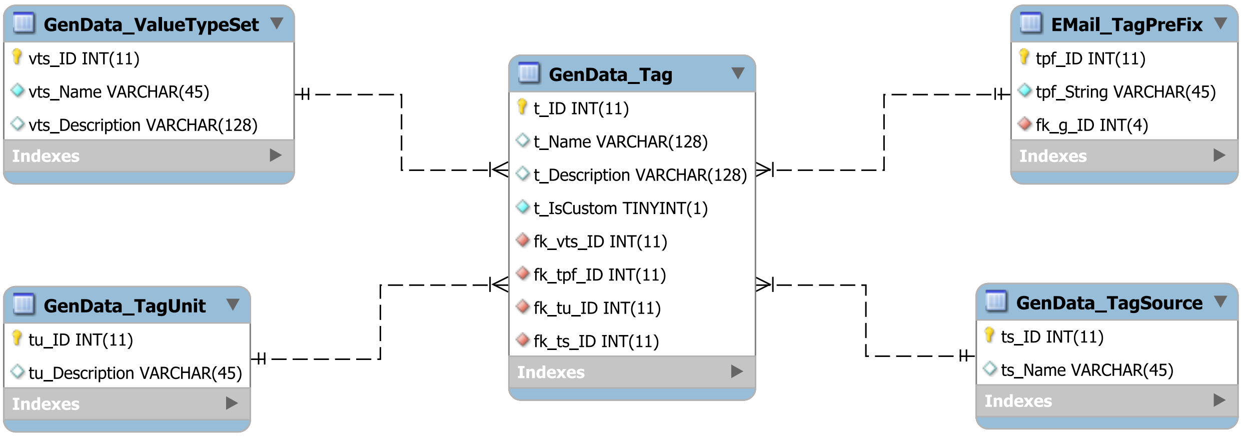


Figure 3.2: *Generic tag structure*

Foreign keys are used to link identifying fields from other tables with the tag. This enables the system or administrators to select pre-existing values and assign it to the tag. Powerful advantages of referential databases are obtained through denormalised tables. Valid fields are created and maintained through referential integrity. Single entry modifications allow bulk updates to be applied to all linked tables.

3.3.3 Storing processed values

During processing data are extracted from files and stored in the database. Refer to Figure 3.3. The *GenData_ValueProcessed* table as shown in the figure is used to store the processed data. The primary key or unique identifier of the *GenData_ValueProcessed* table is a composite key derived by combining the record date with a *vt_ID* from the

GenData_ValueType table, as well as *t_ID* from the *GenData_Tag* table. A value is assigned to each row and will be used to generate reports.

The *GenData_ValueType* table contains a list of known value sample types. Typical examples are time labels such as *01:00* or total values such as *daily_total*. Value sets are formed by combining entries contained in *GenData_ValueType*. *GenData_ValueTypeSet* provides a list of sets. Each set is linked to a number of entries in the *GanData_ValueType* table through the linking table, *GenData_ValueTypeLINK_ValueTypeSet*.

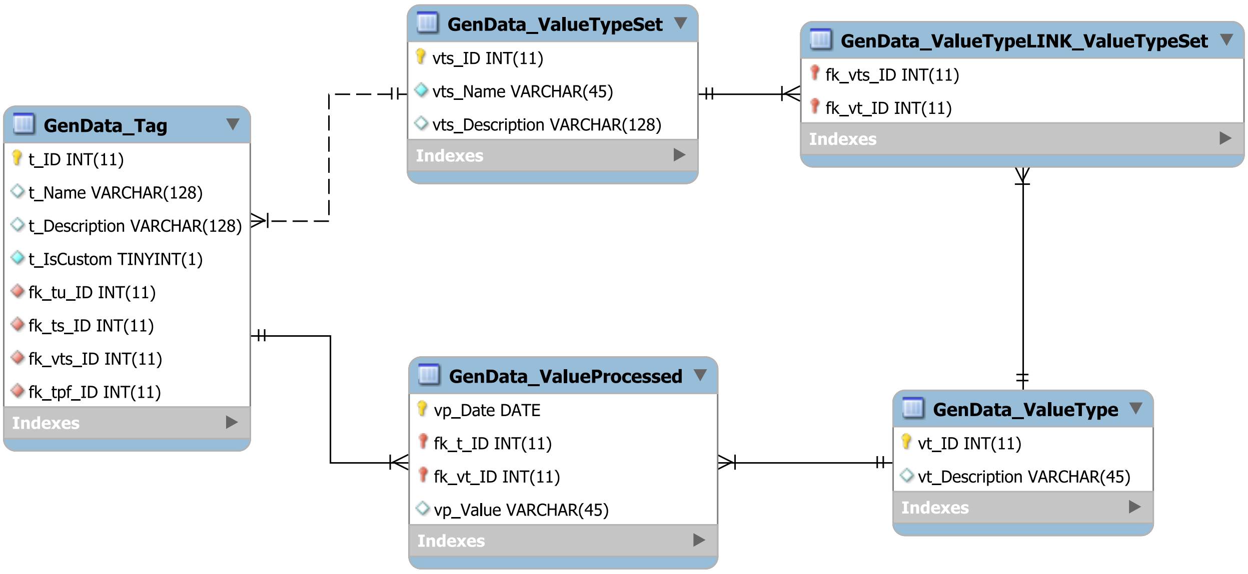


Figure 3.3: *Data storage structure*

Tags contained within the *GenData_Tag* table use are *fk_vts_ID* foreign key column to link to items in the *GenData_ValueTypeSet* table. The linking system discussed previously is used to find value types linked to a specific tag. For each value type assigned to the tag, a composite key consisting of the date, *t_ID* and *vt_ID* is derived. The corresponding value is extracted from the input file and stored with the derived key in the *GenData_ValueProcessed* table.

3.3.4 E-mail handling structure

The collection of *E-Mail* tables as shown Figure 3.4 are used to link e-mail messages and their contents to processes. The file system used to manage the processing procedure is created in correspondence with the structure provided by this group of tables.

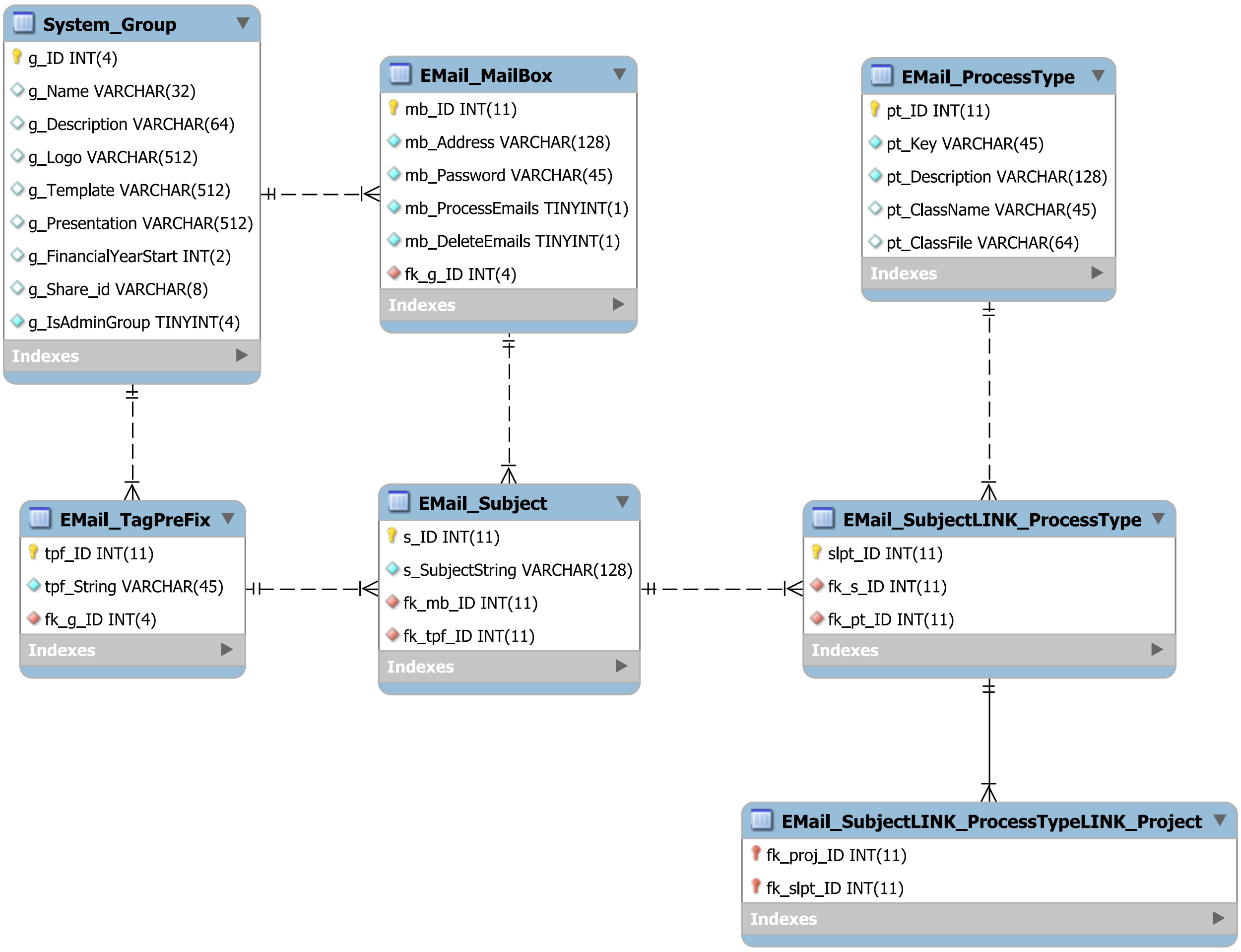


Figure 3.4: *E-mail handling tables*

The *EMail_Mailbox* table contains a list of mailboxes and their respective details. A private mailbox is assigned to each organisation. These organisations are typically mining groups or industrial groups and are stored in the *System_Group* table. Groups are linked to items in the *EMail_TagPreFix* table as well. E-mail subjects are used to identify specific messages in linked mailboxes. Each subject is linked to a specific *tpf_ID* with a foreign key to for navigational purposes.

E-mail subjects are linked to process types contained in the *EMail_ProcessType* table. The *Email_SubjectLINK_ProcessType* table is used to create this link. In cases were the data are used for more than one purpose, *EMail_SubjectLINK_ProcessTypeLINK_Project* uses *slpt_ID* to link received messages to external projects .

A list of existing processes are stored in the *EMail_ProcessType* table. The table provides a description of each entry, along with the class name and file in which it is contained. During processing linked processes will be executed on linked files that are stored in the file system. A discussion of the physical file structure is provided later in this chapter.

3.3.5 Importing REMS files

REMS (Real-time Energy Management System) log files are some of the primary data sources of this system. REMS was developed by the same company that use the system developed in this study. Files created by REMS will thus be used frequently by the developed system and is considered as one of the primary data sources.

The REMS process has a special set of tables used to describe files. Consider Figure 3.5. *EMail_ProcessType* and *EMail_SubjectLINK_ProcessType* tables have been discussed in previous sections. In the case of *REMS* files, four auxiliary tables are used to manage the data import process.

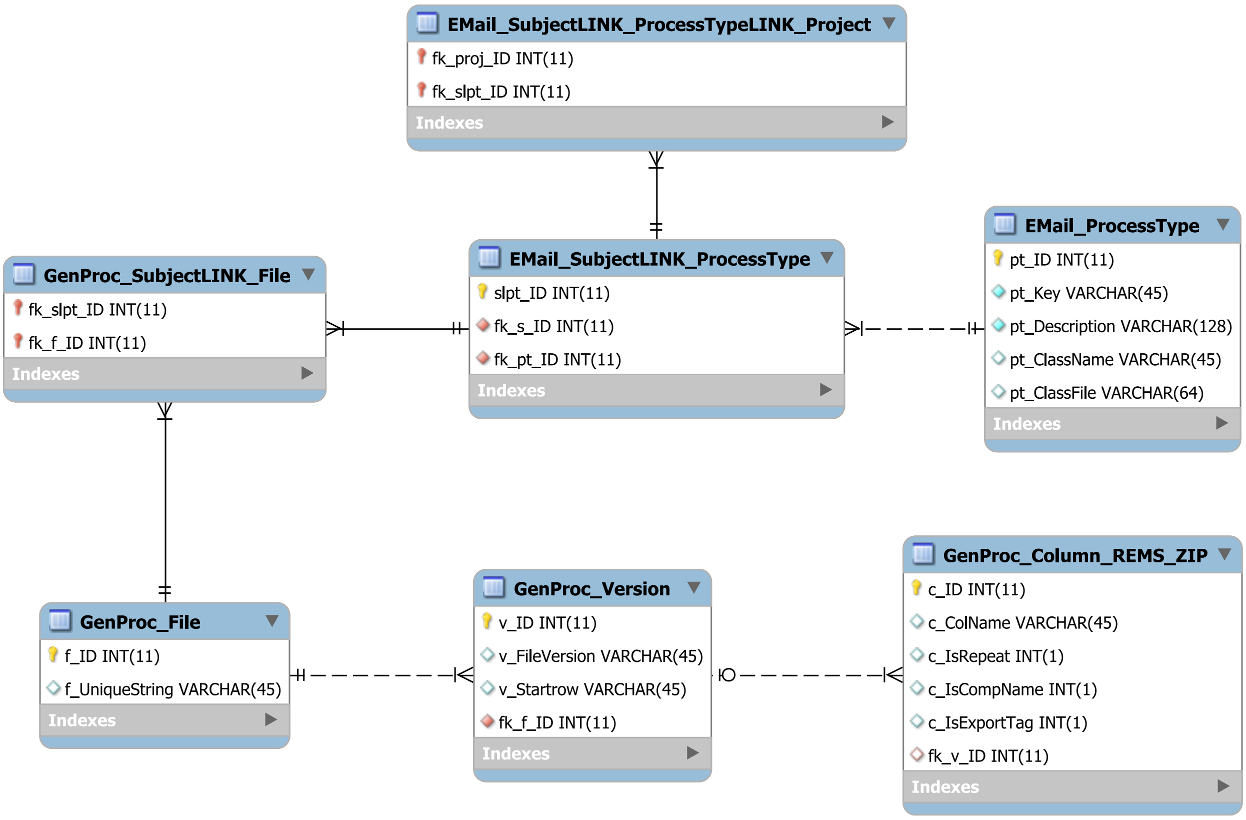


Figure 3.5: *REMS file structure*

GenProc_File is used to store a list of possible REMS files. Generic identification phrases are identified and stored within the table. Unique phrases in file names are linked to downloaded e-mails and used to select appropriate files from a list of possibilities. The file version is linked to a specific column configuration specified in the *GenProc_Column_REMS_ZIP*. *GenProc_Version* is used to create links between files and column configurations.

3.3.6 PDF import database structure

The final import facility developed to date allows data import from intermediate database tables. An external data collection application is used to extract data from PDF (Portable Document Format) files such as accounts. The external application accessed the database and stores collected results in the *GenData_PdfImport* table. Specified values are stored along with a range of identifiers.

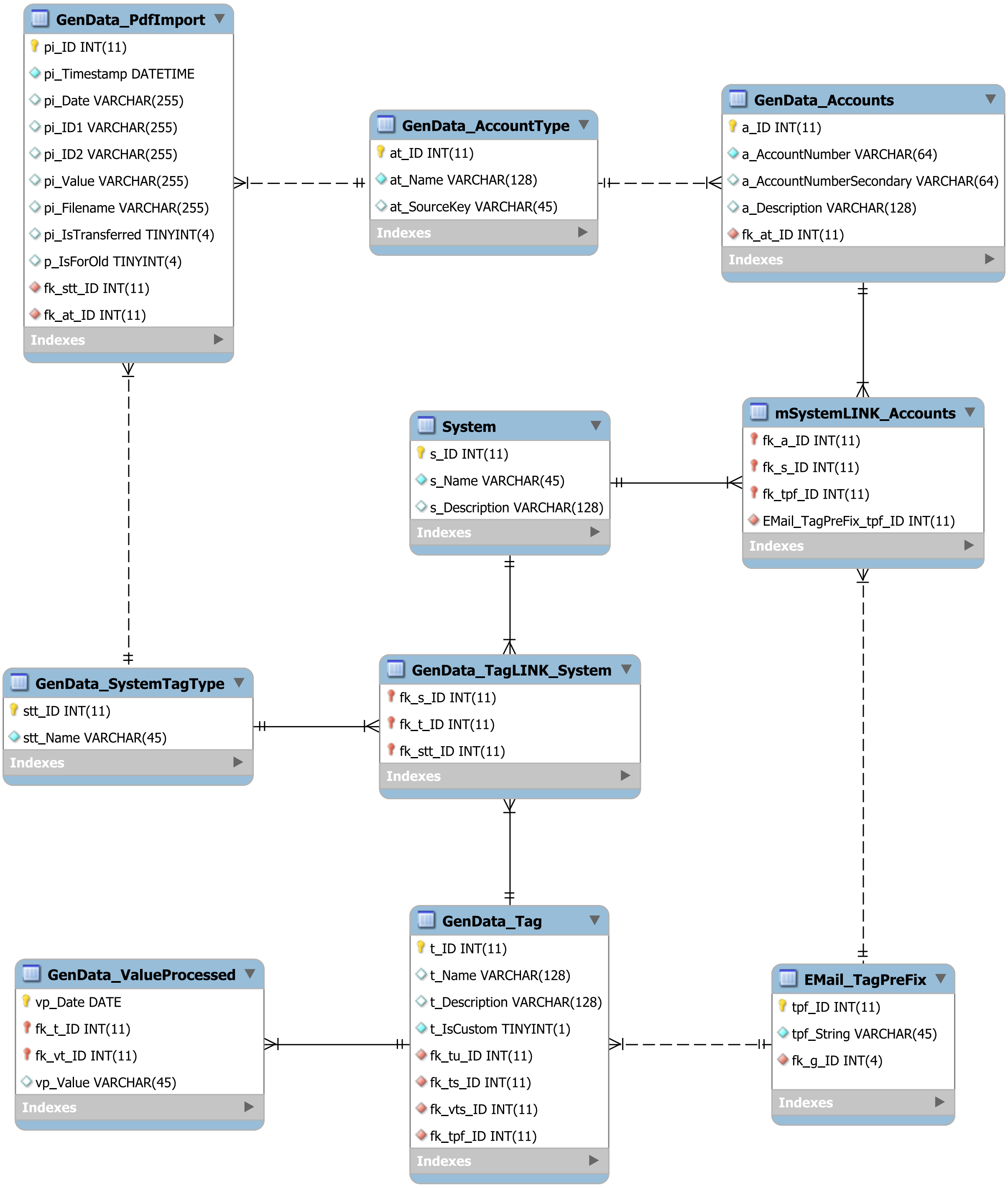


Figure 3.6: *PDF import database tables*

Figure 3.6 provides a visual representation of the tables used to import PDF data. The shown supporting tables are used to link the imported values to specific systems and accounts. Tags can be created using support data linked to the imported values. The data is then transferred to the *GenData_ValueProcessed* table and stored for reporting purposes.

3.3.7 Event log structure

Figure 3.7 shows the event log structure used by the developed system. The *EventLog* table serves as origin for the structure. A record is created in the *EventLog* table for every significant event. *EventLog_Type* stores a list of possible event types. Every possible event type has a unique identifier supplemented with additional details.

Application specific tables are linked to *EventLog* using its primary key as identifier. *Eventlog _EMailTransmit* and *Eventlog_EMailAttachment* are used to store additional information associated with incoming and outgoing e-mail messages. *Eventlog_EMailAttachment* uses a reference from *EMail_Subject*. This reference is used to track specific recorded events.

Eventlog_ProcessFile and *Eventlog_ProcessFileLINK_ValueProcessed* are linked to unique events recorded in *Eventlog*. *Eventlog_ProcessFile* contains details associated with imported files and their stored location. The links indicated in the figure connects a number of tables to *Eventlog_ProcessFileLINK_ValueProcessed*.

The network of tables linked to *Eventlog_ProcessFileLINK_ValueProcessed* is used to track processing events. Each entry in the aforementioned table consists of a unique identifier, a tag reference, a file reference and a data date reference. This enables system administrators to track data changes and restore unwanted changes from files that are stored in the archive.

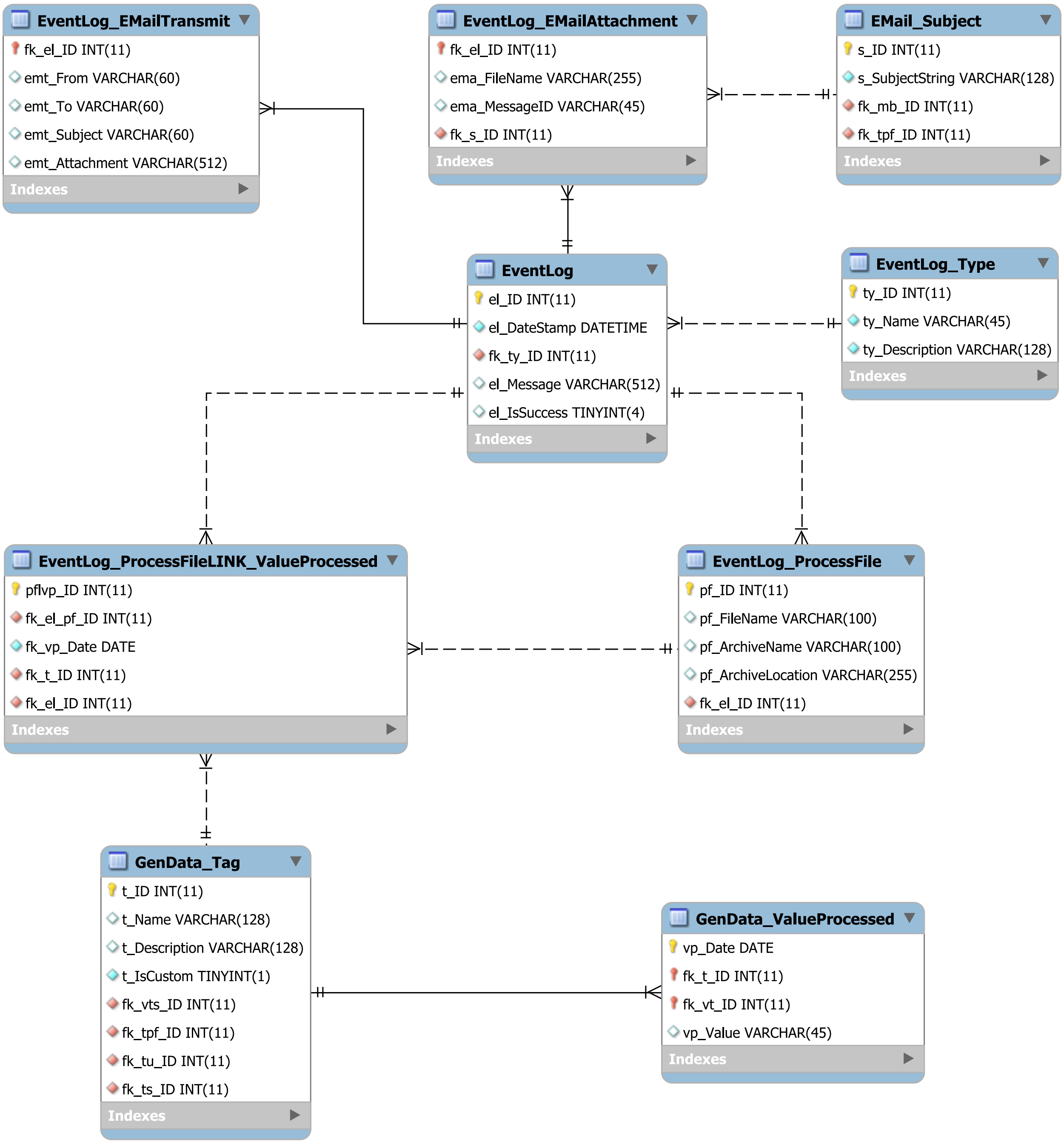


Figure 3.7: *Event log structure*

3.4 Developed software modules

3.4.1 Overview of the process

The database structures discussed in the preceding section are used for two main purposes during processing. Firstly, the structures are used to determine the system's structure. This includes details such as which e-mail mailboxes to check, where files should be stored once downloaded and what processes should be executed on specific tasks. Secondly, processed results are stored in the corresponding database tables.

Processing takes place on a scheduled basis. A list of steps are followed to complete the processing procedure. Figure 3.8 shows the operational flow used to process generic data. The steps indicated in the figure provide a high level description of the process. Each of the steps will be discussed in more detail throughout the remainder of this section.

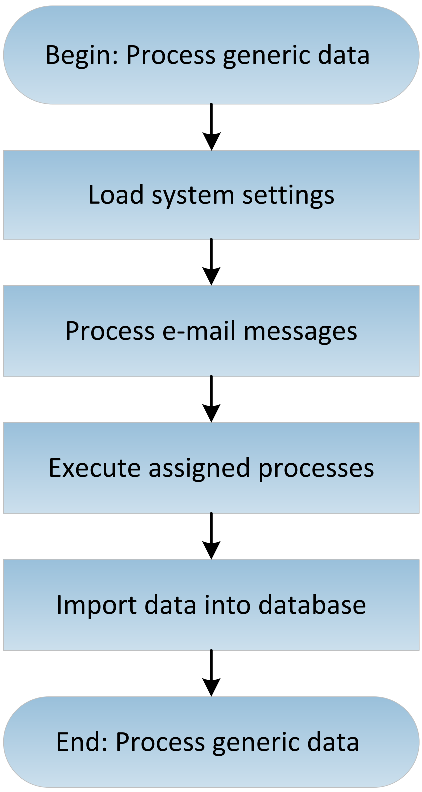


Figure 3.8: *Operational flow – Process generic data*

The basic process, as shown in the figure, cycles through the indicated steps. First, system settings are extracted from the database. Once the settings are loaded from the database it is used to perform the other processes. Next assigned e-mail mailboxes are checked for new data. In the event where data is received it is moved to physical storage for further processing steps.

Unique processes are assigned to specific files, which are then executed in order to convert files to the *Generic* format. The created *Generic* files are moved to import folders and temporarily stored for the import process. Next, the files are imported into the database and archived for safekeeping.

Event logs are created and updated throughout the entire processing procedure. This can be used to trace faults and retrieve historic data. The remainder of this section will be used to discuss the individual processing steps in more detail.

The loaded settings are used to create a physical file structure. The file structure is used as a workspace while performing processing functions. First it is used to store temporary files during processing. After processing is completed, files are archived for safekeeping purposes.

3.4.2 Loading system settings

The system depends on e-mail mailboxes as primary data source. Mailboxes are used to create a structure based on the assigned client groups. Figure 3.9 shows the operational flow used to load system settings. The system iterates through the list of mailboxes and gathers all the necessary information needed to access data.

Received data are classified using e-mail subjects. A list of e-mail subjects are assigned to each mailbox. During each mailbox iteration phase the system compiles a list of linked e-mail subjects. Additional information such as processes linked to the e-mail messages and physical file structure information are compiled using another iterative process.

The process is complete once the system has created the list of all the known mailboxes, subjects and assigned links. This list is used as a reference by the other process steps. Received data that is not linked in this list will be ignored until process links are defined.

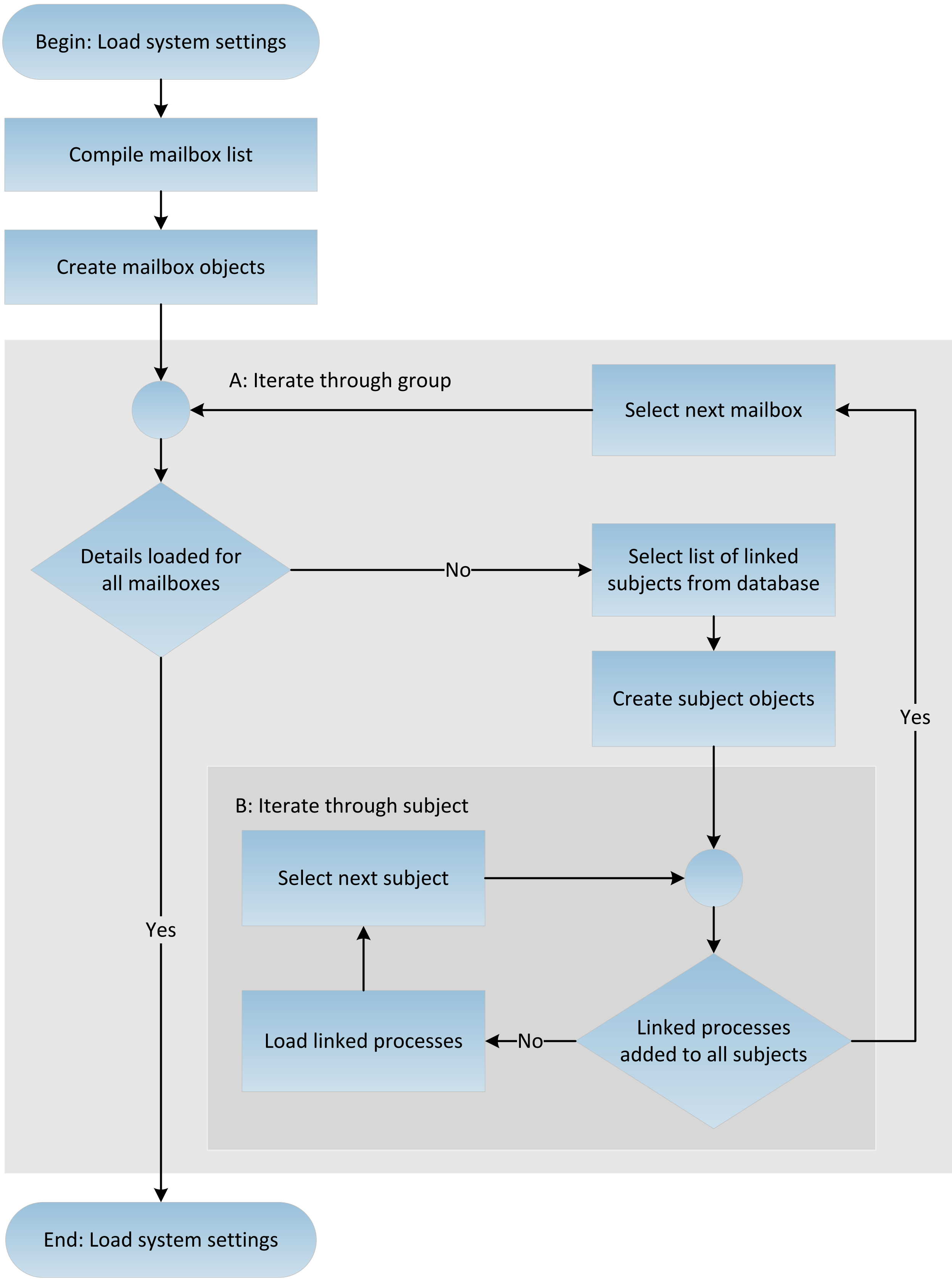


Figure 3.9: *Operational flow – Load system settings*

3.4.3 Processing e-mail messages

Figure 3.11 shows the process followed when processing e-mail messages. The loaded system settings discussed previously are used as references. Known mailboxes marked for processing are selected. The system iterates through the list of linked mailboxes in order to collect received data.

While cycling through the list of mailboxes, the system iterates though a list of linked subjects. Each subject is linked to a specific process. Figure 3.10 shows an example of the file structure used during processing. The system downloads the data contained within the e-mail messages and stores it in an assigned *_PENDING* folder under the corresponding *Group* and *Prefix* directories.

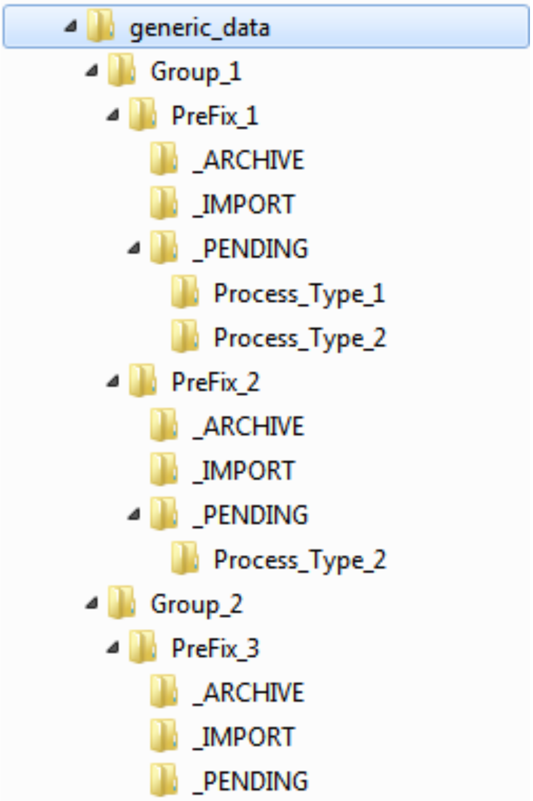


Figure 3.10: *Physical file structure example*

If messages are marked for deletion, they are removed from the mailbox after the download is complete. The system continues with this process until all the known mailboxes and linked processes have been processed. Unknown messages will be ignored until a process is defined. If no action is required the message must be removed by a system administrator after an acceptable time period.

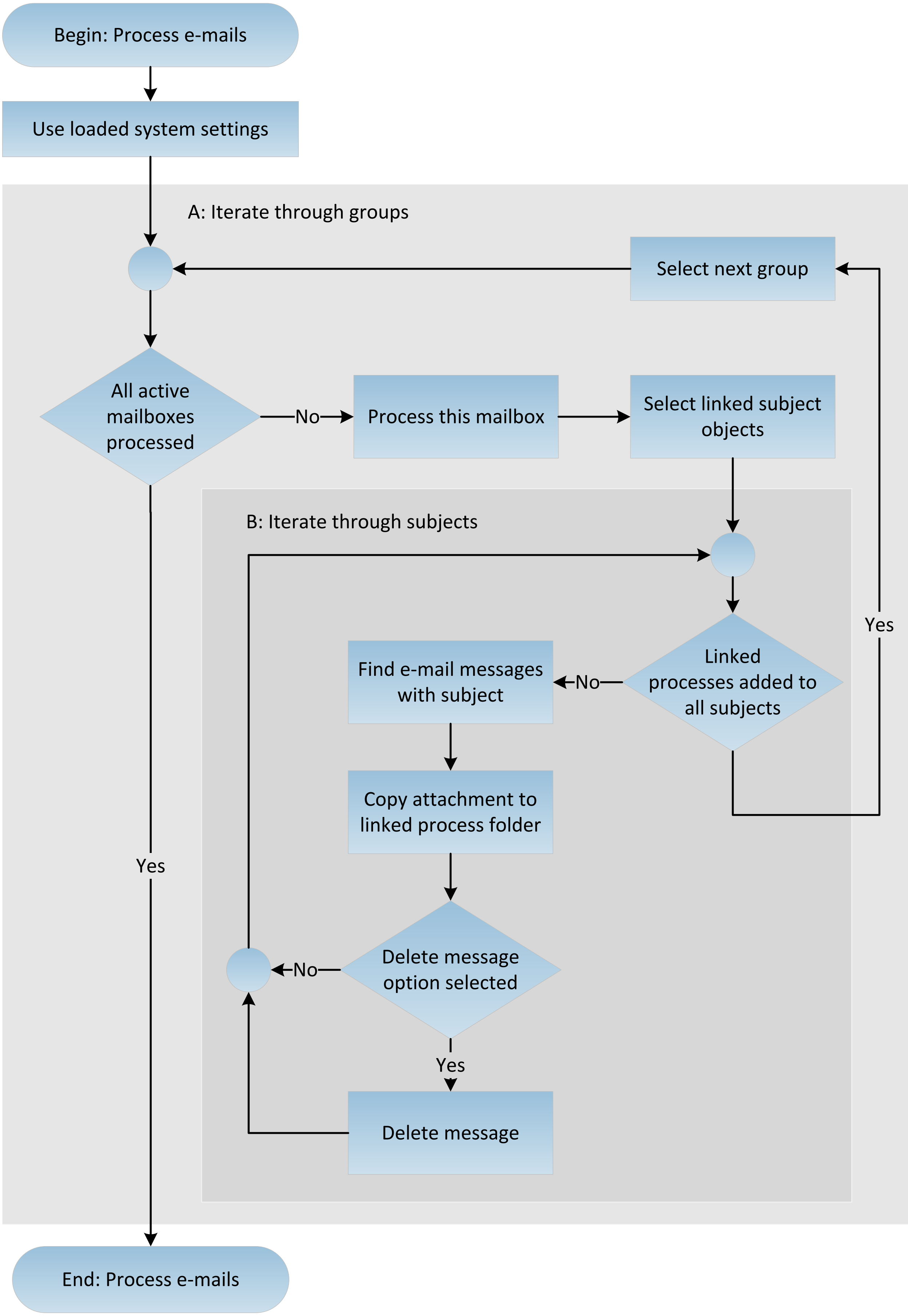


Figure 3.11: *Operational flow – Process e-mail messages*

3.4.4 Processing pending data

The primary objective of the process pending file process is to convert data from multiple formats to a single generic data format. Figure 3.12 shows a graphical representation of the process. The process executes in a iterative fashion similar to the processes discussed previously.

Preloaded system settings are used as a starting point. The settings are used as a reference when searching for files. Stored files pending processing are arranged according to groups, sub-groups or prefixes, and process folders within the *_PENDING* folders. Refer to Figure 3.10 to see the physical file structure.

The system iterates through the group directories and linked subdirectories. These subdirectories are linked to e-mail subjects and determines which *PreFix* folder to consider. Files waiting to be processed are stored within the *_PENDING* folder. Inside the *_PENDING* folder files are classified and stored in sub-directories, that are linked to possible process types.

While iterating through the various *_PENDING* folders the system calls assigned processes. Each process can be seen as a black box that expects a particular input and provides a specific output. The input file format determines which data extraction process is required. During processing the file is opened and read using fixed parameters.

The read data is restructured and stored in the *Generic* file format. Created *Generic* files are stored in the corresponding *_IMPORT* folders for further processing. In the event where files are received in the *Generic* format, the called extraction process simply moves the files to the corresponding *_IMPORT* folder.

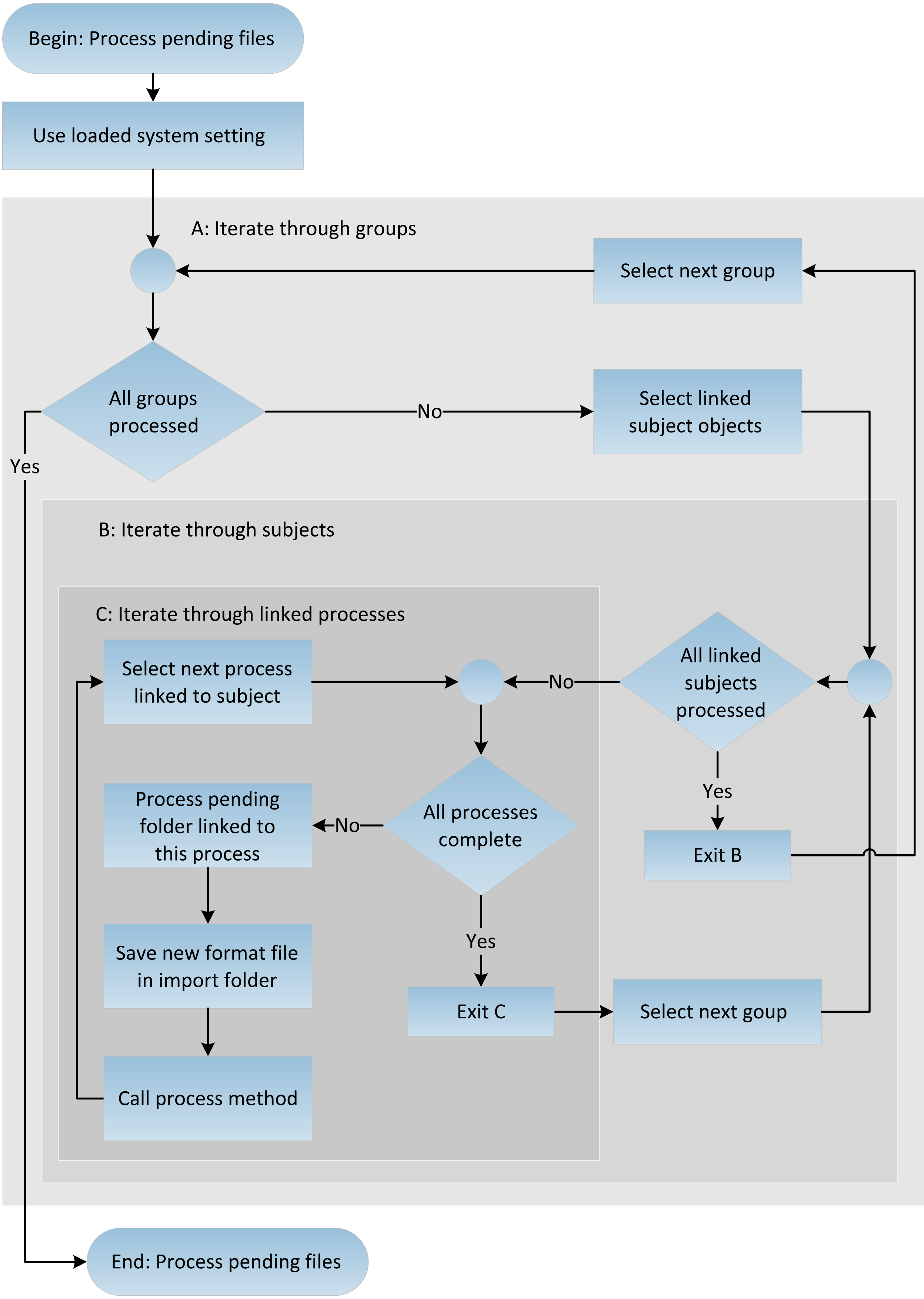


Figure 3.12: *Operational flow – Process pending files*

3.4.5 Importing data to the database

After conversion is complete the main process calls the import function to import data into the database. Only *Generic* format files are considered when importing data. The preprocessing steps discussed previously are responsible to convert the files to this format before data import is possible. Figure 3.14 provides a graphical representation of the import process.

The process includes loading possible *units* and *value type sets* from the database. *Value type sets* are used to select labels for the imported data. These sets are formed by linking a list of known elements such as time intervals, for example, *'00:30, 01:00, 01:30'*, or total samples such as *'monthly_total'*.

Data is read from the *Generic* file by iterating through its rows. While there are unprocessed rows, the system searches for the phrase *Tag Name:* to identify the next data tag. If the tag does not exist the system calls the *Create tag* function to create a new tag in the database. Figure 3.13 provides a graphical representation of the creation process.

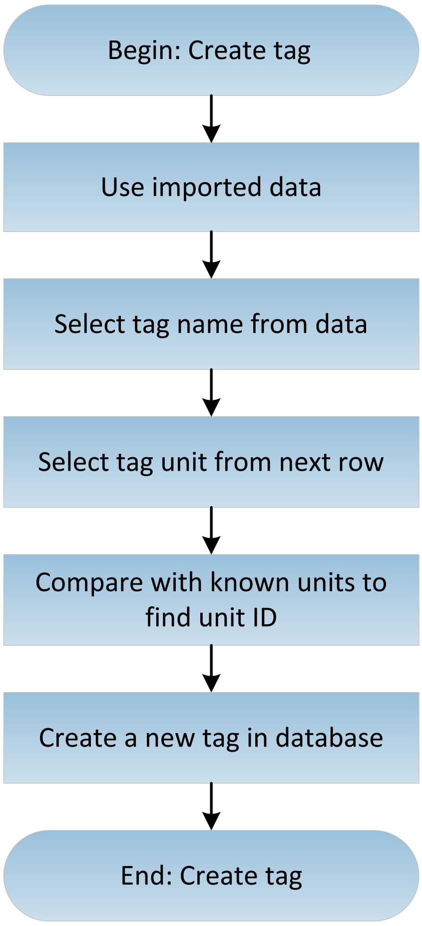


Figure 3.13: *Operational flow – Create tag*

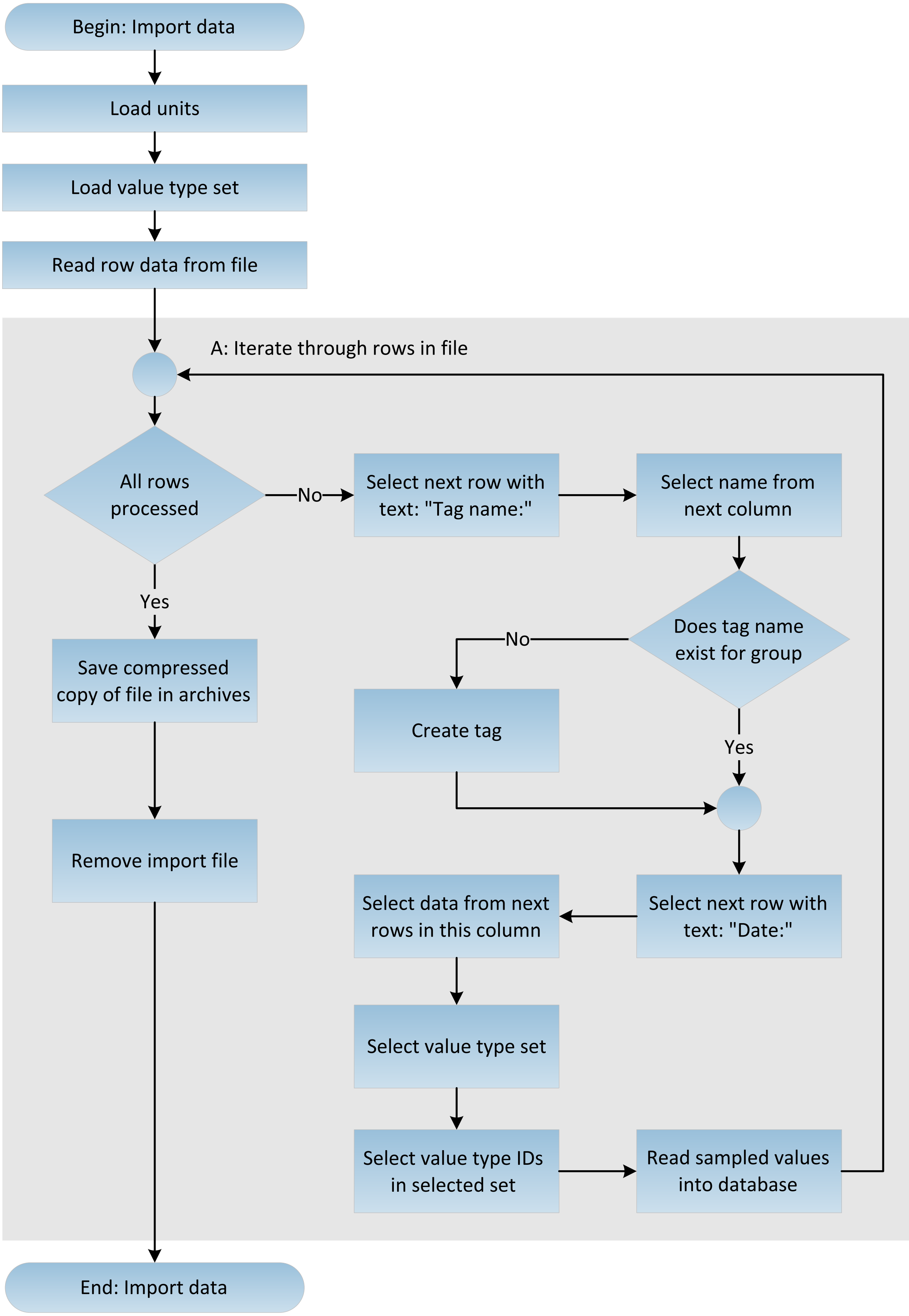


Figure 3.14: *Operational flow – Import data*

Next, sample data is read from the file and classified according to the sample data, as well as corresponding data types. The extracted data are then written to the database, along with event logs for each significant event. SQL (Structured Query Language) functions are used to perform the required database transactions.

Once all the rows have been processed and imported the files are compressed and stored in the corresponding *_ARCHIVE* folder shown in Figure 3.10. If the file has been archived successfully it is removed from the import folder and all temporary folders are removed.

3.5 Consolidating data from files

Processing takes place using the methods described in section 3.4. Unique processes are used to extract data from files of a particular format. Data from the unique file are extracted, rearranged and saved in the *Generic* format. Figure 3.15 shows the *Generic* file structure.

Tag ID:	C_1 ID			
Tag Name:	C_1 Name			
Date:	YYYY-MM-DD	YYYY-MM-DD	···	YYYY-MM-DD
Parameter 1	$C_1D_1P_1$	$C_1D_2P_1$	···	$C_1D_nP_1$
Parameter 2	$C_1D_1P_2$	$C_1D_2P_2$	···	$C_1D_nP_2$
⋮	⋮	⋮		⋮
Parameter n	$C_1D_1P_n$	$C_1D_2P_n$	···	$C_1D_nP_n$
Tag ID:	C_2 ID			
Tag Name:	C_2 Name			
Date:	YYYY-MM-DD	YYYY-MM-DD	···	YYYY-MM-DD
Parameter 1	$C_2D_1P_1$	$C_2D_2P_1$	···	$C_2D_nP_1$
Parameter 2	$C_2D_1P_2$	$C_2D_2P_2$	···	$C_2D_nP_2$
⋮	⋮	⋮		⋮
Parameter n	$C_2D_1P_n$	$C_2D_2P_n$	···	$C_2D_nP_n$
⋮	⋮	⋮		⋮
Tag ID:	C_n ID			
Tag Name:	C_n Name			
Date:	YYYY-MM-DD	YYYY-MM-DD	···	YYYY-MM-DD
Parameter 1	$C_nD_1P_1$	$C_nD_2P_1$	···	$C_nD_nP_1$
Parameter 2	$C_nD_1P_2$	$C_nD_2P_2$	···	$C_nD_nP_2$
⋮	⋮	⋮		⋮
Parameter n	$C_nD_1P_n$	$C_nD_2P_n$	···	$C_nD_nP_n$

Figure 3.15: *File structure – Generic*

The reformatted file is saved in the group's import folder. If received files are already in the *Generic* file format, the base process will simply move the file to the group's import folder. Three file formats are commonly used by the system. The *Generic* data format is the most versatile format and is used as the primary system format. All other files are converted to this format before it is imported.

Figure 3.16 shows the *IST* file structure. This type of file is obtained by exporting data from Powertech (2014) databases. Figure 3.17 shows *REMS* format and how the file is structured. Because REMS files are created by software programs, developed by the same group, that fund the data consolidation platform, a set of tables have been developed specifically to extract data from this type of file.

Date Time	Component 1	Component 2	···	Component n
YYYY-MM-DD 00:30:00	T_1C_1	T_1C_2	···	T_1C_n
YYYY-MM-DD 01:00:00	T_2C_1	T_2C_2	···	T_2C_n
⋮	⋮	⋮		⋮
YYYY-MM-DD 24:00:00	T_nC_1	T_nC_2	···	T_nC_n

Figure 3.16: *File structure – IST*

Time	Mode	P_1	P_2	···	P_n	P_1	P_2	···	P_n	···	P_1	P_2	···	P_n
T_1	M_x	$C_1T_1P_1$	$C_1T_1P_2$	···	$C_1T_1P_n$	$C_2T_1P_1$	$C_2T_1P_2$	···	$C_2T_1P_n$	···	$C_nT_1P_1$	$C_nT_1P_2$	···	$C_nT_1P_n$
T_2	M_x	$C_1T_2P_1$	$C_1T_2P_2$	···	$C_1T_2P_n$	$C_2T_2P_1$	$C_2T_2P_2$	···	$C_2T_2P_n$	···	$C_nT_2P_1$	$C_nT_2P_2$	···	$C_nT_2P_n$
⋮	⋮	⋮	⋮		⋮	⋮	⋮		⋮		⋮	⋮		⋮
T_n	M_x	$C_1T_nP_1$	$C_1T_nP_2$	···	$C_1T_nP_n$	$C_2T_nP_1$	$C_2T_nP_2$	···	$C_2T_nP_n$	···	$C_nT_nP_1$	$C_nT_nP_2$	···	$C_nT_nP_n$

Figure 3.17: *File structure – REMS*

3.6 System integration

The database components and software component that were described in the preceding sections were integrated with the existing energy management system. The existing system provides the necessary framework that is used to generate reports for specific purposes. Additionally, the existing system acts as a graphical interface used to display results after processing the data.

General system constraints (as described previously) eased the integration process. New functionality was added by integrating the components of the study with the existing system. Development of auxiliary components such as data transfer channels and hosting services were avoided through integration. The combined system addresses all the set requirements.

3.7 Development summary

In this chapter the development of the data consolidation platform was discussed. Both the developed database components, as well as software components were described. File handling was discussed at the hand of actual files. Benefits of integration with the existing system were obtained by conforming to set system constraints.

The developed structures offer fundamental facilities to prove that the system is capable of the required outcomes. After integration, the system was tested and improved on a iterative basis. The resulting system is fully generic and offers users a wide range of options to collect and manage energy related data. The next chapter will focus on the results obtained by using the developed platform.

4

Evaluation of the data consolidation platform

The platform was developed as discussed in the preceding chapters. After implementation the database was populated using sample data from clients in a range of industries. The developed platform was used to consolidate data into a single comprehensive data set. Results of the consolidation process are used to evaluate the efficacy of the platform.

4.1 Evaluation introduction

Chapter 3 discussed the development of the data consolidation platform. The platform was developed in alignment with the needs and constraints discussed in Chapter 1 and Chapter 2. In this chapter, the platform is evaluated to determine its efficacy.

Data from seven large energy consumers in the South African industry were consolidated using the developed system. The data originated from a range of sources and were stored in an assortment of files with unique formats. The system was used to consolidate these files and store the results in a generic format.

In the remainder of this chapter the platform will be assessed. Data import results will be discussed in detail. Thereafter, data extraction results will be assessed along with examples. The platform will be evaluated to determine its correlation with the specifications. Lastly a summary of the results will be provided.

4.2 Data consolidation results

4.2.1 Data file import

Data were collected by a range of external systems and stored in each system's unique format. The collected data was received by the system as attachments in e-mail messages. The data was downloaded and stored in a logical file structure. These files are submitted to predefined processing steps and imported into the generic database.

Specific modules are used to extract and restructure the data. Additional modules can be added to the system without modifying the core system. The file contents were successfully converted to the *Generic* format and consolidated with the database. Storing data in a single format improves the robustness of the system and offers users a single file standard. Figure 4.1 show an example of a *Generic* file.

Both *IST* and *REMS* formatted files were successfully converted to the *Generic* format. Figure 4.2 shows an example of a *REMS* formatted file. The converted files are stored and imported during the import phase of the data consolidation process, as discussed in section 3.4.1. Figure 4.3 shows an example of a *IST* file.

Tag ID:	-1						
Database:	ION						
Tag Name:	ION Power						
Date:	2014-07-20	2014-07-21	2014-07-22	# #	2014-08-17	2014-08-18	2014-08-19
01_00	6486	4494	6138	# #	5836	6016	6082
02_00	6377	6225	5044	# #	5861	6050	6099
03_00	4815	6221	4454	# #	5869	6062	5982
04_00	4662	4433	4383	# #	5824	5385	5957
05_00	5867	4432	4401	# #	4921	4462	5957
06_00	6320	5718	4479	# #	4254	4426	6095
07_00	5566	6193	2941	# #	4269	4341	6043
08_00	4223	3934	2615	# #	5640	4346	6063
09_00	4217	3269	2630	# #	5976	3946	5697
10_00	4241	4404	2572	# #	5980	251	2527
11_00	5282	4384	4767	# #	6017	2793	2519
12_00	6189	5703	6175	# #	5995	3994	5334
13_00	6304	6198	6148	# #	5976	4441	6059
14_00	6333	6201	6116	# #	6040	5556	5611
15_00	6327	5045	6105	# #	6029	6005	2632
16_00	6262	4451	4482	# #	6023	6007	4045
17_00	6306	6029	3641	# #	5979	6086	5637
18_00	6315	4457	6147	# #	5991	5607	5096
19_00	6351	397	1030	# #	4838	364	320
20_00	6372	380	317	# #	4308	323	296
21_00	6450	3492	2704	# #	4316	4196	3241
22_00	6345	6181	6175	# #	5742	6100	5916
23_00	5382	6164	5696	# #	6056	5988	5896
24_00	4817	6123	4499	# #	6048	5971	5950

Figure 4.1: *File example – Generic*

Date	Time	Platform mode	Name	Write permission	Air enthalpy	Air dry bulb temp	Air relative humidity	Air ambient pressure	Enthalpy Min	Enthalpy Max
2014-10-10	0:00:00	Auto	Weather Station	0	?	28.52	24.61	88	0	100
2014-10-10	0:02:00	Auto	Weather Station	0	?	28.64	23.77	88	0	100
2014-10-10	0:04:00	Auto	Weather Station	0	?	28.62	22.93	88	0	100
2014-10-10	0:06:00	Auto	Weather Station	0	?	28.73	23.17	88	0	100
2014-10-10	0:08:00	Auto	Weather Station	0	?	28.73	23.62	88	0	100
2014-10-10	0:10:00	Auto	Weather Station	0	?	28.65	23.85	88	0	100
2014-10-10	0:12:00	Auto	Weather Station	0	?	28.56	23.98	88	0	100
2014-10-10	0:14:00	Auto	Weather Station	0	?	28.5	24.04	88	0	100
2014-10-10	0:16:00	Auto	Weather Station	0	?	28.28	23.66	88	0	100
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
2014-10-10	23:40:00	Auto	Weather Station	0	?	23.44	53.95	88	0	100
2014-10-10	23:42:00	Auto	Weather Station	0	?	23.38	53.68	88	0	100
2014-10-10	23:44:00	Auto	Weather Station	0	?	23.39	54.32	88	0	100
2014-10-10	23:46:00	Auto	Weather Station	0	?	23.51	55.12	88	0	100
2014-10-10	23:48:00	Auto	Weather Station	0	?	23.42	53.76	88	0	100
2014-10-10	23:50:00	Auto	Weather Station	0	?	23.35	53.35	88	0	100
2014-10-10	23:52:00	Auto	Weather Station	0	?	23.38	54.18	88	0	100
2014-10-10	23:54:00	Auto	Weather Station	0	?	23.34	54.63	88	0	100
2014-10-10	23:56:00	Auto	Weather Station	0	?	23.31	54.11	88	0	100
2014-10-10	23:58:00	Auto	Weather Station	0	?	23.22	54.92	88	0	100

Figure 4.2: *File example – REMS*

Date Time	Incomer1_Pnet	Incomer2_Pnet	# # # # # # # # # # #	Motor6_Pnet	Overhead Line_Pnet
2014-09-11 0:30	1152	348	# # # # # # # # # # #	246	204.0000153
2014-09-11 1:00	1152	336	# # # # # # # # # # #	246	192.0000153
2014-09-11 1:30	1120	324	# # # # # # # # # # #	244.7999878	192.0000153
2014-09-11 2:00	1104	300	# # # # # # # # # # #	246	174
2014-09-11 2:30	1120	300	# # # # # # # # # # #	246	180
⋮	⋮	⋮	⋮	⋮	⋮
2014-09-11 21:30	1376	348	# # # # # # # # # # #	144	192.0000153
2014-09-11 22:00	1328	252	# # # # # # # # # # #	0	156
2014-09-11 22:30	1360	384.0000305	# # # # # # # # # # #	0	215.9999847
2014-09-11 23:00	160	60	# # # # # # # # # # #	100.7999954	53.99999619
2014-09-11 23:30	784	276	# # # # # # # # # # #	246	120
2014-09-12 0:00	1440	431.9999695	# # # # # # # # # # #	246	180

Figure 4.3: *File example – IST*

4.2.2 Document import results

An external tool is used to extract data from PDF files. Figure 4.4 and 4.5 show the interface of the PDF manager application. The application is used to select known fields in a PDF document based on position. This enables users to select fields such as account numbers, names and values.

The selected values are imported using the database table structure discussed in section 3.3.6. Values extracted from PDF files are temporarily stored in these tables and transferred to processed tables during processing. Values extracted from bills can be used in cases where metering equipment is not available. Alternatively, these values can be compared with calculated values for verification purposes.

A test case was used to verify the accuracy of the data consolidation platform with an Eskom account. Total electricity consumption value is selected from the Eskom account, as shown in Figure 4.5. The amount reflected on the bill is 3 721 474.70 kWh.

Electricity consumption values were measured by the client at the facility. The values were recorded and imported by the data consolidation platform. A monthly report was generated to compare the results. The calculated total electricity consumption is 3 634 434 kWh. The calculated value only shows a slight variation in accuracy and can be attributed to calibration differences, line losses and other electrical losses. Figure 4.6 shows the calculated results.

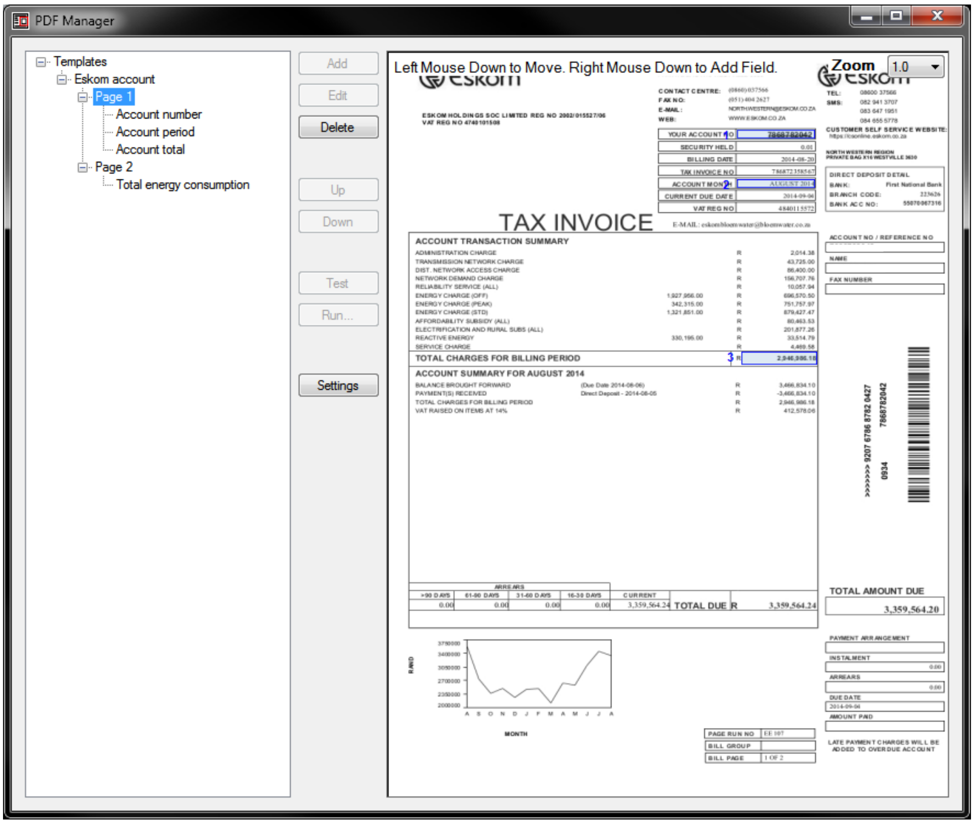


Figure 4.4: *PDF manager application window 1*

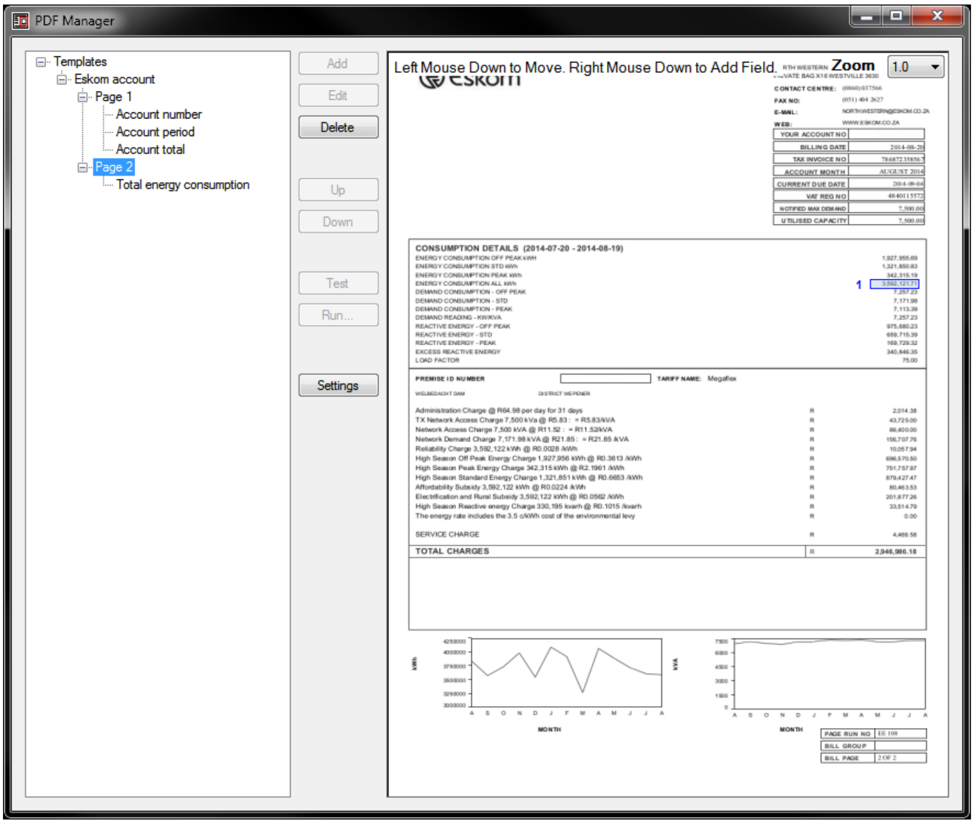


Figure 4.5: *PDF manager application window 2*

Daily performance tracking

Table 5-1: System summary for the period

Date	Actual usage	Budget usage		Actual cost	Budget cost	
2014-07-20	137 809 kWh	133 185 kWh		R 56 199	R 54 313	
2014-07-21	114 528 kWh	135 554 kWh		R 92 366	R 134 388	
2014-07-22	103 659 kWh	135 554 kWh		R 80 259	R 134 388	
2014-07-23	127 759 kWh	135 554 kWh		R 111 966	R 134 388	
2014-07-24	101 550 kWh	135 554 kWh		R 84 584	R 134 388	
2014-07-25	124 817 kWh	135 554 kWh		R 103 420	R 134 388	
2014-07-26	117 974 kWh	132 899 kWh		R 60 166	R 67 172	
2014-07-27	116 789 kWh	133 185 kWh		R 47 627	R 54 313	
2014-07-28	83 507 kWh	135 554 kWh		R 80 902	R 134 388	
2014-07-29	90 815 kWh	135 554 kWh		R 90 302	R 134 388	
2014-07-30	133 934 kWh	135 554 kWh		R 124 328	R 134 388	
2014-07-31	135 705 kWh	135 554 kWh		R 120 947	R 134 388	
2014-08-01	47 607 kWh	135 554 kWh		R 40 419	R 134 388	
2014-08-02	138 300 kWh	132 899 kWh		R 71 415	R 67 172	
2014-08-03	138 684 kWh	133 185 kWh		R 56 555	R 54 313	
2014-08-04	129 513 kWh	135 554 kWh		R 113 609	R 134 388	
2014-08-05	124 721 kWh	135 554 kWh		R 98 131	R 134 388	
2014-08-06	116 610 kWh	135 554 kWh		R 95 742	R 134 388	
2014-08-07	131 201 kWh	135 554 kWh		R 115 334	R 134 388	
2014-08-08	121 637 kWh	135 554 kWh		R 98 893	R 134 388	
2014-08-09	108 172 kWh	132 899 kWh		R 52 585	R 67 172	
2014-08-10	128 627 kWh	133 185 kWh		R 52 454	R 54 313	
2014-08-11	114 736 kWh	135 554 kWh		R 108 770	R 134 388	
2014-08-12	98 218 kWh	135 554 kWh		R 93 042	R 134 388	
2014-08-13	107 686 kWh	135 554 kWh		R 88 445	R 134 388	
2014-08-14	121 578 kWh	135 554 kWh		R 109 276	R 134 388	
2014-08-15	123 792 kWh	135 554 kWh		R 108 429	R 134 388	
2014-08-16	136 948 kWh	132 899 kWh		R 69 504	R 67 172	
2014-08-17	133 788 kWh	133 185 kWh		R 54 559	R 54 313	
2014-08-18	108 716 kWh	135 554 kWh		R 82 363	R 134 388	
2014-08-19	115 054 kWh	135 554 kWh		R 95 670	R 134 388	
Total:	3 634 434 kWh	4 179 707 kWh		R 2 658 261	R 3 496 789	

Legend

	Within budget
	Over budget
	Weekends
	No verified data available

Figure 4.6: *Result – Monthly calculation*

4.3 Using consolidated data for reporting purposes

The developed data consolidation platform was combined with an existing web-based energy management system. The existing system provides a graphical interface and components used to access and display stored data. After integrating the platform with the existing system the combination was used to construct a variety of energy reports.

The energy management platform has the ability to convert between units. A unit is assigned to each imported tag. These units are used to convert or normalise data when generating reports. Auxiliary tags are used as conversion variables. Measurement data from other tags are multiplied with conversion values of the auxiliary tags in order to produce results in specific formats.

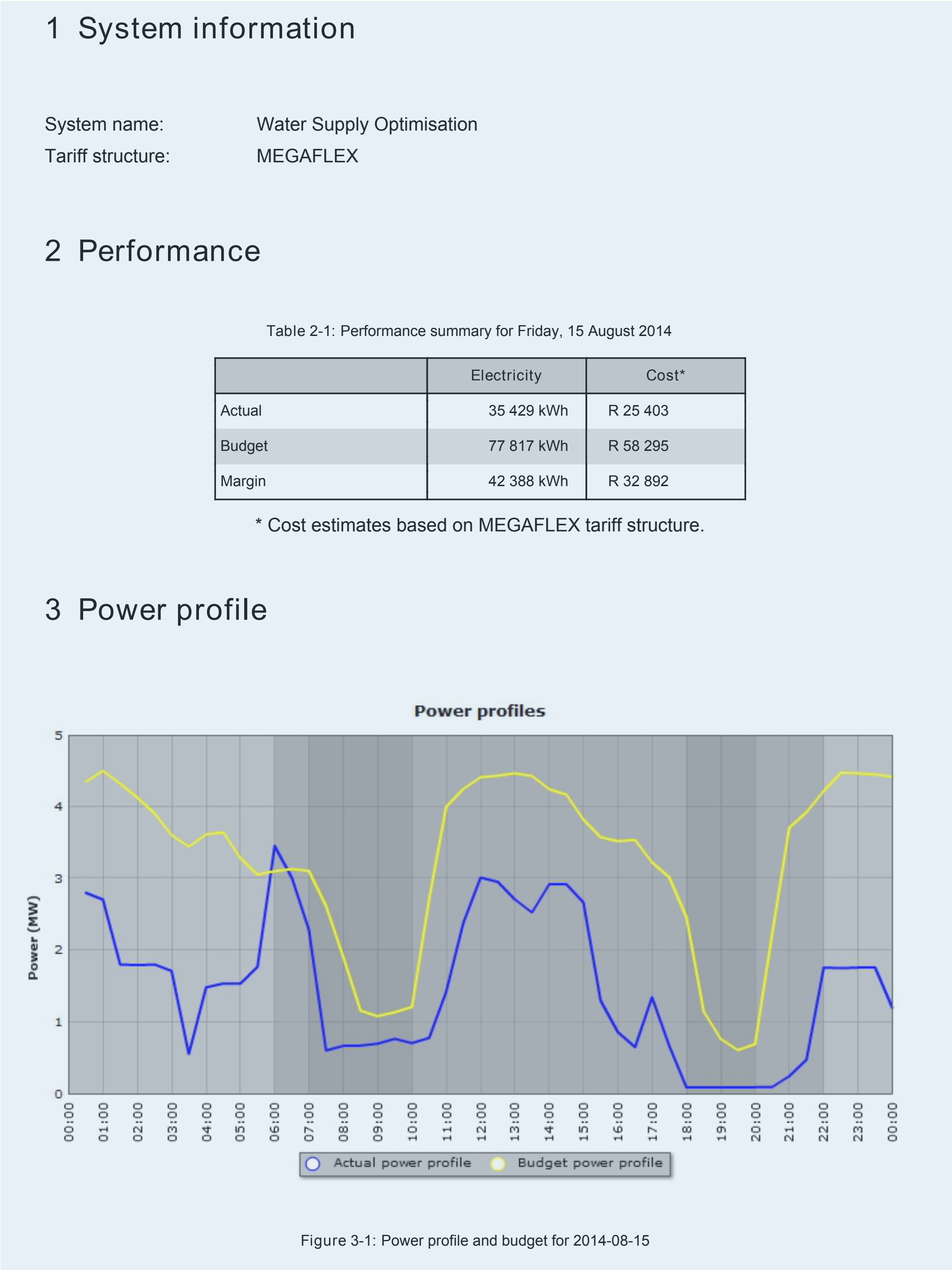
Four applications are presented in this section. The applications are used to present the platform's capabilities by showing the different reports generated by the integrated system. Data requirements are unique for all the presented outcomes. The platform's unique generic data handling capabilities allows it to import and store the wide range of data that is shown.

The purpose of the presented applications is to validate the data consolidation platform. The aim of the platform is to consolidate energy consumption data and production data into a single data set. Information is then extracted from the data set in order to compile energy reports. These reports satisfy the energy reporting needs of the South African market, as discussed in section 1.4.1.

Application 1: DSM Project reporting

The integrated system was used to construct a report used to monitor the performance of a DSM project. To achieve this, an existing DSM project was used as a test subject. Performance data was exported from the client's database. A separate application is used to establish a ODBC (Open Database Connectivity) connection to the database, access the data and store it in the *Generic* format, as seen in section 3.5.

The data consolidation platform was used to import the data. Next, the corresponding tags and their values were used to create a daily performance report. In this case only electricity data was taken into account. Figure 4.7 shows the report generated by the integrated system.



1 System information

System name: Water Supply Optimisation
Tariff structure: MEGAFLEX

2 Performance

Table 2-1: Performance summary for Friday, 15 August 2014

	Electricity	Cost*
Actual	35 429 kWh	R 25 403
Budget	77 817 kWh	R 58 295
Margin	42 388 kWh	R 32 892

* Cost estimates based on MEGAFLEX tariff structure.

3 Power profile

Figure 3-1: Power profile and budget for 2014-08-15

Figure 4.7: *Result – Integrated system*

The graph shows the actual energy consumption in blue, compared with the budgeted energy consumption in yellow. The baseline found in the M&V report was used as the budget. The calculations show a 42 388 kWh saving margin for for the 24-hour period, which translates to a cost saving of R 32 892.00. The cost saving was calculated based on Eskom's MEGAFLEX tariff structure.

The results were verified using two methods. Firstly, the actual energy consumption values, as well as M&V baseline values were entered into Microsoft Excel®. The data sets consist of 48 samples per profile, per day (one sample every 30 minutes). The difference between the actual and budgeted samples were calculated and summed over the 24-hour period to determine the combined electricity saving for the day.

Hourly cost values were calculated by multiplying hourly energy consumption figures with corresponding tariff values. As mentioned earlier, Eskom's Megaflex tariff structure was used to perform these calculations. The calculated results corresponded with the values calculated by the integrated system, thus proving the system's accuracy.

As a secondary control, the results of the integrated system were compared to the system developed by Goosen. Figure 4.8 shows the results obtained after entering the same raw files into Goosen's system. Again, the yellow line indicates the baseline power, while the blue line indicates actual power values.

The resulting graphs exhibit roughly the same profile as the developed system. The key difference is that the developed system presents a higher resolution output with 48 increments for the 24-hour period, while Goosen's system only shows 24 increments. The developed system has the capability to display even higher data resolutions if higher accuracy is required.

Cost values calculated by Goosen's system correspond with the values calculated by the developed system. Both systems indicate a cost saving of R32 892.00 over the 24-hour period. This further proves the developed system's accuracy and validates the results calculated by the system.

Monetary savings are calculated using hourly energy consumption figures. Identical data sets were imported into both systems discussed in this section. The calculated saving amounts are exactly the same, due to the identical data sets. This resulted in equal hourly energy consumption totals, as well as equal hourly cost figures.

1 Project information

Project name:	Water Supply Optimisation
Project number:	2011-05
Tariff structure:	MEGAFLEX
Target impact:	1.51 MW
Target energy saving:	36.24 MWh per day

2 Performance (Friday 2014-08-15)

Performance of day:		Month-to-date performance:	
Impact:	1.77 MW	Average impact:	2.14 MW
Energy saving:	42.39 MWh	Cumulative energy saving:	564.86 MWh
Cost saving:	R 32 892	Cumulative cost savings:	R 552 325
Missed opportunities:	R 2 045	Cumulative missed opportunities:	-

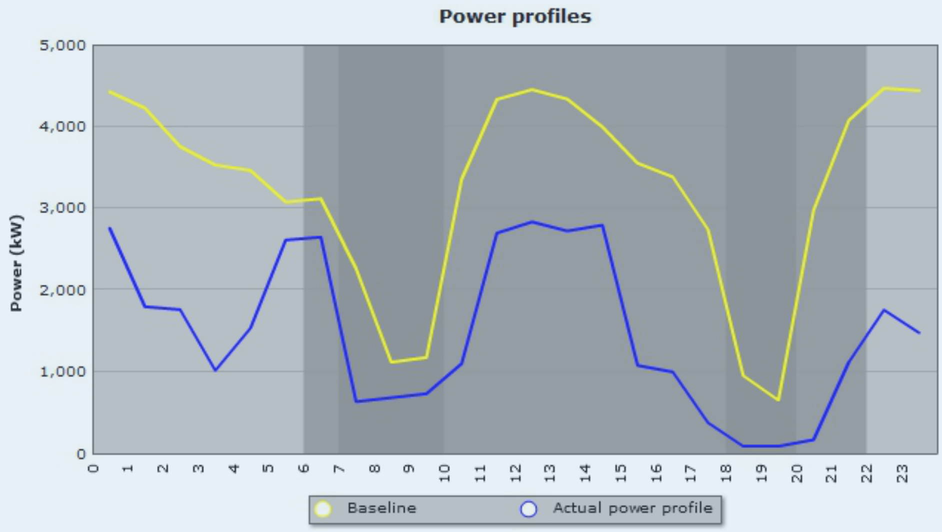


Figure 2-1: Power profile and baseline for 15 August 2014

Figure 4.8: *Result – Goosen's system*

Application 2: Tax incentive reporting

Another key outcome of the generic data consolidation system is the ability to store a range of data formats. Figure 4.9 shows an example of a report generated with the aim to participate in the tax incentive programme discussed in section 1.4.4. The table as shown in the figure shows that various energy data types are accommodated by the system.

A unit type is assigned to tags to provide this generic capability. Processed values for these tags are stored according to dates and assigned sample intervals. These stored values were made accessible by integrating the data consolidation platform with the energy management system.

Summary

Table 1: Summary

Energy source	2012 total	2013 total	GWh saving	Tax incentive
Gas GJ	750 510	648 846	294	R 37 057 595
Electricity (kWh)	322 167 556	305 293 501	145	R 18 278 926
Coal (Tons)	10 074 894	9 505 229	1 969	R 218 359 222
Total			2 408	R 273 695 743

Total production

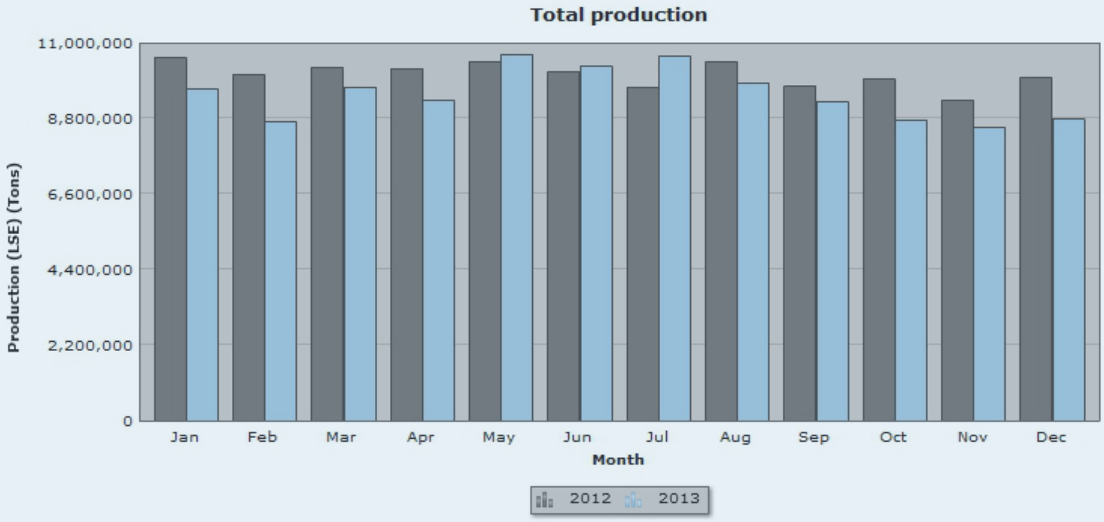


Figure 1-1: Total production

Figure 4.9: *Result – Tax incentive report*

As mentioned previously, energy savings obtained across all energy types are considered when calculating tax rebates according to section 12L of the income tax act. Various energy sources and figures related to these specific elements can be combined through the use of the platform. Consumers can assess the feasibility of interventions by regarding the energy sources in a single summary.

Application 3: Mandatory provision of energy data reporting

According to changes proposed to the National Energy Act of South Africa, large energy consumers will be forced to disclose energy data. The developed data consolidation platform offers a solution to energy consumers and can be used as a tool to gather the required energy data. Figure 4.10 shows the draft report template as provided by the Department of Energy.

QUESTIONNAIRE 4-23			
Sub-Sector	**MINING OF GOLD**		
SIC Code	**23**	**Version**	**Nov-2013-v1**

1. RESPONDENT DETAILS
Please provide details in the table below of the facility for which this response is being made:

Name of firm	
Name of facility	
Contact details	Name: E-mail: Telephone:

2. REPORTING PERIOD
Please specify which twelve-month period this response relates to:

Calendar Year [ ]

3. ENERGY CONSUMPTION
Please enter the total energy consumption of the facility, over the time period specified in Question 2, for each of the energy carriers in the table below:

Energy carrier	**Consumption**	**Units**
Electricity		kWh/ MWh/ GWh
Coal		Tonnes
Natural gas		m3
Off-grid renewable		please specify units
Petroleum products:		
Diesel		Thousands litres
Petrol/gasoline		Thousands litres
Paraffin/kerosene		Thousands litres
LPG		Thousands litres
Light fuel oil		Thousands litres
Heavy fuel oil		Thousands litres
Other		Thousands litres
Total Energy Consumption		**TJ**

4. PHYSICAL OUTPUT
Please indicate in the table below the production level of the facility over the time period specified in Question 2:

Product	Quantity	Units
Ore mined Final product		thousand tonnes kg

Figure 4.10: *Resource reporting template*

Demonstration reports were generated after integrating the data consolidation platform with the energy management system. Because of the high data sensitivity actual data could not be used. Fictive data was used to demonstrate system capabilities. The test data provides adequate proof that the platform is capable of storing all the required data elements to produce these reports.

Data was entered in the form of monthly totals. In cases such as oil and coal, monthly averages were added to provide supplementary information. In many cases petroleum fuels received are recorded in litres. However, energy values are calculated based on Joule per tonne calculations. To allow the conversion intermediate variables are used to estimate liquid density in the form of kilogram per litre.

The energy management system has the capability to perform basic calculations on data. The product of the three input variables mentioned previously results in total energy figure in the form of Joule. Thus, energy units can be converted to other forms by multiplying stored values with conversion values in order to produce the results in the form required by the DoE.

Table 1: Questionnaire for the period: 2012-01-01 to 2012-12-31

Questionnaire	2-34241
Sub-sector	Manufacture of X
SIC Code	34241
Version	Nov-2013-v1

Table 2: Respondent details for the period: 2012-01-01 to 2012-12-31

Details		
Consumer details	Firm name:	Demo firm
	Facility name:	Facility x
Contact details	Name:	Employee
	E-mail:	email@demo.com
	Telephone:	012-345-6789

Table 3: Reporting period for the period: 2012-01-01 to 2012-12-31

Calendar year:	2012

Figure 4.11: *Result – Resource report part A*

Figure 4.11 and Figure 4.12 show an equivalent report as produced by the integrated energy management system. The results as shown are year totals of the stored monthly totals. Specified units are used to report on the various energy sources, in addition to the calculated total energy consumption that is provided in Joule as specified.

Table 4: Energy consumption in X for the period: 2012-01-01 to 2012-12-31

Energy carrier	Consumption	Units
Electricity	140 238	MWh
Coal	182 759	Tonnes
Natural gas	0	m3
Off-grid renewables	0	x
Diesel	1 509	Thousand litres
Perol	56	Thousand litres
Paraffin	0	Thousand litres
LPG	1 159	Thousand litres
Light fuel oil	15	Thousand litres
Heavy fuel oil	0	Thousand litres
Total:	8	TJ

Table 5: Physical output for the period: 2012-01-01 to 2012-12-31

Product	Quantity	Units
X	100 000	Tonnes

Figure 4.12: *Result – Resource report part B*

Application 4: Performance tracking

The platform offers users the ability to create reports using data from several sources. The graph in Figure 4.13 shows energy consumption values per hour compared to ambient temperature. From the graph a clear correlation can be seen between the hourly temperature average and energy consumption of the cooling system.

Temperature values were recorded by a weather station and stored as a *REMS* file. Power data was recorded and stored in *IST* format by a different system on the same facility. However, these values were not available on the same system.

Measured values from both systems were compared after consolidating the measured files into a single database using the developed platform. This enables users to create advanced performance tracking reports. Without the data consolidation platform the measured values will not be available into a single system and manual calculations will be required to produce a similar report.

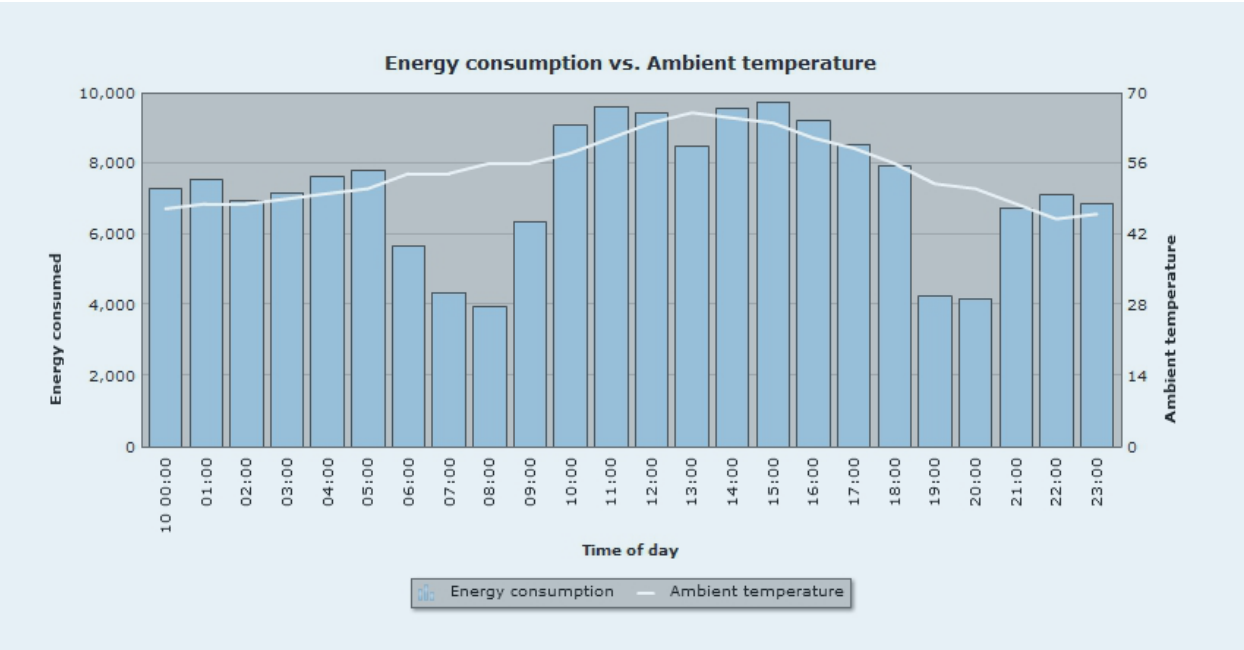


Figure 4.13: *Result – Performance report*

4.4 Evaluation of specification compliance

Table 4.1 shows the list of specifications as discussed in section 3.2.2. The specifications are divided into four classes namely: general, Demand Side Management reporting, tax incentive and resource reporting. Individual requirements are provided next to the class indicators, along with an indication whether the system addresses the specific requirement. From the table it can be observed that the system addresses all the required outcomes.

Table 4.1: *Addressed specifications*

Class	Specification	Addressed
General	The platform should accommodate energy data.	Yes
	The system must be capable of storing production data.	Yes
	Data sampled at fixed intervals must be accommodated.	Yes
	Budget and actual data must be stored for comparison.	Yes
	Original files must be archived for M&V purposes.	Yes
	The platform must accommodate a range of data sources.	Yes
	The platform must be capable of accommodating new processing extensions without modifications to its framework.	Yes
DSM	The platform must be able to store electricity data.	Yes
	The platform must accept electricity data in the form of energy data and power data.	Yes
	Data samples for numerous intervals during a single day should be accommodated.	Yes
	Budget and tariff data must accompany measured data and should be handled.	Yes
	Relevant production and climate data should be stored for support purposes.	Yes
Tax incentive	The platform must be able to store energy data.	Yes
	Energy data is not limited exclusively to electricity, but includes all energy data types.	Yes
	Energy baselines, budget data and actual measurements must be stored.	Yes
	Data is not limited exclusively to electricity, but includes all energy data types.	Yes
	Various data intervals should be accommodated, such as monthly and daily figures.	Yes
Resource reporting	Actual energy data and production data must be handled.	Yes
	Data is not limited exclusively to electricity, but includes all energy data types.	Yes
	Various data intervals should be accommodated, such as monthly and daily figures.	Yes
	The system should be able to handle various measurement units.	Yes
	The platform must be capable of consolidating the measurements to a single quantifiable total.	Yes

4.5 Summary of results

This section provides proof that the developed platform meets the set of specifications. The platform offers a generic platform to import and store data from various inputs. Data units and sample intervals are taken into account. Auxiliary information can be imported in parallel with measurement data.

Reporting options were made available by integrating the system with the existing energy management system. Reports are generated by extracting data from the database. The generic database structure offers users the capability to generate unique reports based specific needs.

The platform has the ability to adapt so that new file formats can be imported. To show this three different file examples were shown. The respective file types were all imported using the system. During the import process the system was capable of creating tags, extracting associated data from the files and loading the data into a relational database.

Data was extracted from an account document using an external application. The application stored results in a set of tables. Data stored in these tables were used to create tags and store the extracted data in the database. The values were compared with results calculated by the system. This proves that the system is capable of calculating accurate totals over large time periods.

Custom reports were constructed using data stored in the database. Four applications were used to show the reporting capabilities of the system. DSM project reporting, tax incentive reporting, mandatory provision of energy data reporting and performance tracking reports were generated. These applications show that the system has the capability to handle a wide range of data, for a wide range of reporting purposes.

Lastly, the specifications discussed in section 3.2.2 were assessed to determine whether the system achieved its goals. Results obtained after practical implementation showed that the system met all the requirements.

5

Conclusion and recommendations

This chapter provides a final discussion of the study. Completed work is summarised and discussed. Furthermore, a conclusion of the study is provided, followed by recommendations for further development in the field of data consolidation.

5.1 Summary of work completed

Globally there is a need for improved energy management. Governments around the world offer a range grants of for energy efficiency projects. Improved data consolidation methods are needed to satisfy reporting requirements related to these grants. In South Africa alone, two legislative acts have been adapted and require improved energy data management and energy reporting.

Significant changes in the adapted acts of South Africa necessitate the need to report on all forms of consumed energy and not only electricity. In order to do this a new energy data consolidation platform was needed. The data consolidation platform developed in this study is capable of consolidating energy data of various data types in a single database.

The developed system has the capability to extract data from multiple sources and store the data in a single generic format. New modules can be added in order to allow specialised data extraction from unique files. These specialised software modules can be added to the system without major changes to the core framework. This allows users to expand the system on a modular basis.

After development of the data consolidation platform, the system was integrated with an existing web-based energy management system. The integrated energy management system offered users the ability to generate a range of reports in alignment with new reporting requirements. If new energy reporting needs arise the system is able to accept the data and will be able generate additional reports with minimal development.

5.2 Study conclusion

A data consolidation platform was researched and developed and was presented as this study. After development was completed the platform was implemented as part of a web-based energy management system. The developed platform addresses all the identified system requirements. Shortfalls associated with related systems were identified and addressed during system development.

An ESCo put the integrated solution to practical use. A wide range of production data with varying data units and sample frequencies were consolidated using the system. Actual client data was successfully imported and used to generate a multitude of energy reports.

Examples of system outputs are shown in the evaluation chapter. The results were verified by manual mathematical analysis, as well as comparison with an existing energy management system. This provided sufficient proof that the system provides accurate results.

Practical use of the system was validated by assessing the systems ability to produce relevant energy reports. Three reporting requirements were identified at the start of the study. This included demand side management reporting, tax incentive reporting, as well as mandatory provision of energy data reporting. Each of the reporting outcomes have unique requirements. The unique requirements were addressed by the system.

In conclusion, the system addressed the set requirements. The data consolidation platform allows users to import and manage a wide range of data from multiple sources. Integration with an existing energy management system allows the generation of various industry specific energy reports. These reports were constructed to meet the requirements of new energy related reports in the South African economy.

5.3 Recommendations for further development

The system was developed to allow generic data consolidation of production data. Various data sources were imported, however the system is capable of achieving more. Restrictions concerning advanced file formats such as *xls* files impede the system from extracting data from these types of files.

Data extraction from documents require development. These documents include accounts or bills that can be used to verify results. However, documents published by external sources are not always constructed according to a fixed template. These changing templates require advanced data extraction tools to identify specific data fields and values and must be developed.

Currently the system only processes data during scheduled intervals. Although the platform is capable of accepting sample data with unique time intervals, typical updates only take place once a day. More regular process intervals will allow near real-time tracking options. This will enable users to detect problems and take proactive management action.

External data sources often cause reliability problems. Direct access to client databases will allow the system to draw marked data on a regular basis and provide up to date results. This will improve the possibility of real time data tracking further, as well as ensure consistency with other client systems.

6

Bibliography

ASHRAE (2013). ANSI/ASHRAE/IES Standard 90.1, Energy Standard for Buildings Except Low-Rise Residential Buildings.

Asif Mushtaque, M., Dhiman, H., Maheshwari, S., and Hussain, S. (2014). Evaluation of DES, TDES, AES, Blowfish and Twofish Encryption Algorithm: Based on Space Complexity. *International Journal of Engineering Research & Technology (IJERT)*.

Blaha, M. (2010). *Patterns of Data Modeling*, volume 1. CRC Press.

Bourgeois, M., Breaux, K., Chiasson, M., and Mauldin, S. (2010). Tax incentives of going green. *CPA Journal*, pages 19 – 24.

Brodkin, J. (2008). Gartner: Seven cloud-computing security risks. *Infoworld*, pages 1–3.

CarbonetiX (2013). Smart Portal^{TM}, http://www.carbonetix.com.au/products/.

Chaudhari, M. and Patel, S. (2014). A Survey on Cryptography Algorithms. *International Journal of Advance Research in Computer Science and Management Studies*.

Chen, P. (1976). The entity-relationship modeltoward a unified view of data. *ACM Transactions on Database Systems (TODS)*, pages 9–36.

Convery, S. (2004). *Network security architectures*. Pearson Education India.

Copin, L., Rey, H., Vasques, X., Laurent, A., and Teisseire, M. (2010). Intelligent energy data warehouse: What challenges? In *Tools with Artificial Intelligence (ICTAI), 2010 22nd IEEE International Conference on*, volume 2, pages 337–342. IEEE.

Cui, Y. and Widom, J. (2003). Lineage tracing for general data warehouse transformations. *The International Journal on Very Large Data Bases*, pages 41–58.

Den Heijer, W. (2009). The measurement and verification guideline for energy efficiency and demand-side management (EEDSM) projects and programmes. Measurement and verification team, North-West University.

Department of Energy (2010). Policy to support the Energy Efficiency and Demand Side Management Program for the Electricity Sector through the Standard Offer Incentive Scheme. 192 Visagie Street, Pretoria, 0001.

Department of Minerals and Energy (1998). White Paper on the Energy Policy of the Republic of South Africa. Travenna Office Campus, 75 Meintjies Street, Sunnyside, Pretoria, 0001.

Department of Minerals and Energy (2005). Energy Efficiency Strategy of the Republic of South Africa. Travenna Office Campus, 75 Meintjies Street, Sunnyside, Pretoria, 0001.

Department of Public Enterprises (2007). Strategically important developments. 1090 Arcadia Street, Pretoria, 0001.

Efficiency Valuation Organization (2012). International Performance Measurement and Verification Protocol, Concepts and Options for Determining Energy and Water Savings. *Concepts and Options for Determining Energy and Water Savings*, pages 1–143.

El-Sappagh, S., Hendawi, A., and El Bastawissy, A. (2011). A proposed model for data warehouse ETL processes. *Journal of King Saud University - Computer and Information Sciences*, pages 91–104.

Elminaam, D., Abdual-Kader, H., and Hadhoud, M. (2010). Evaluating the performance of symmetric encryption algorithms. *IJ Network Security*, pages 216–222.

Eskom Ltd. (2012). Integrated report for the year ended 31 March 2012. Technical report, Eskom, Megawatt Park, Maxwell Drive, Johannesburg, South Africa.

Eskom Ltd. (2013). COP17 Fact sheet. Technical report, Eskom, Megawatt Park, Maxwell Drive, Johannesburg, South Africa.

Eskom Ltd. (2014). Tariffs and Charges. Technical report, Eskom, Megawatt Park, Maxwell Drive, Johannesburg, South Africa.

Eskom Media Desk (2014). Eskom calls on all customers to reduce electricity usage, as it declares an emergency. Technical report, Eskom, Megawatt Park, Maxwell Drive, Johannesburg, South Africa.

Eyadat, M. and Larsen, S. (2011). A new heuristic approach to safeguarding online digital data. *Eighth International Conference on Information Technology: New Generations*, pages 1–4.

Felden, C. and Chamoni, P. (2003). Web farming and data warehousing for energy tradefloors. In *Web Intelligence, 2003. WI 2003. Proceedings. IEEE/WIC International Conference on*, pages 642–645. IEEE.

Fiedler, T. and Mircea, P. (2012). Energy management systems according to the ISO 50001 standard challenges and benefits. In *Applied and Theoretical Electricity (ICATE), 2012 International Conference*, pages 1–4.

Finley, M. (2013). BP Statistical Review of World Energy. Technical report, BP Technical Report.

Fowler, M. (2004). *UML distilled: a brief guide to the standard object modeling language.* Addison-Wesley Professional.

Frank, L. (2011). Countermeasures against consistency anomalies in distributed integrated databases with relaxed acid properties. In *Innovations in Information Technology (IIT), 2011 International Conference on*, pages 266–270. IEEE.

Gillis, M. (2009). PowerLogic EEM Energy Management Information System (EMIS). Technical report, Schneider Electric, 1 Riverview Office Park, Janadel Avenue, Johannesburg, South Africa.

Gökçe, H., Wang, Y., Gökçe, K., and Menzel, K. (2009). A data-warehouse architecture supporting energy management of buildings. In *Einhoven, Netherlands*. CIB W78.

Goosen, P. (2013). Efficient monitoring of mine compressed air savings. Master's thesis, Faculty of Engineering, North-West University.

Gouchoe, S., Everette, V., and Haynes, R. (2002). Case Studies on the Effectiveness of State Financial Incentives for Renewable Energy. In *National Renewable Energy Laboratory*, pages 50 – 55. North Carolina State University.

Gruschka, N. and Jensen, M. (2010). Attack surfaces: A taxonomy for attacks on cloud services. In *Cloud Computing (CLOUD), 2010 IEEE 3rd International Conference on*, pages 276–279. IEEE.

Hamer, W. and Vosloo, J.C. and Swanepoel, J.A. (2013). Electricity costs and the financial feasibility of low grade gold ore processing. In *Industrial and Commercial Use of Energy Conference (ICUE), 2013 Proceedings of the 10th*, pages 1–6. IEEE.

Hirani, S. (2008). Energy Consumption of Encryption Schemes in Wireless Devices. Master's thesis, Faculty of Engineering, University of Pittsburgh.

Hoang, D., Tran, H., Nguyen, B., and Tjoa, A. (2012). Towards the development of large-scale data warehouse application frameworks. In *Re-conceptualizing Enterprise Information Systems*, pages 92–104. Springer.

Hooke, J., Hart, D., and Landry, B. (2004). *Energy Management Information Systems: Achieving Improved Energy Efficiency: A Handbook for Managers, Engineers and Operational Staff.* Office of Energy Efficiency of Natural Resources Canada.

House of Representatives of the United States of America (2005). Energy Policy Act of 2005.

Inmon, W. (2005). *Building the data warehouse.* John Wiley & Sons, United States of America, 3 edition.

International Organization for Standardization (2011). ISO 50001 energy management. *ISO Central Secretariat Chemin de Blandonnet 8 CP 401 1214 Vernier, Geneva Switzerland*, pages 4–7.

Jain, Y. and Gosavi, P. (2008). Email Security Using Encrption and Compression. In *Computational Intelligence for Modelling Control & Automation, 2008 International Conference on*, pages 136–139. IEEE.

Kimball, R. and Caserta, J. (1996). The data warehouse. *ETL Toolkit. Practical Techniques for. Extracting, Cleaning,. Conforming, and. Delivering Data*.

Levi, A. and Ko, K. (2011). Inside risks: Risks in email security. *Communications of the ACM*, pages 112–114.

Lloyd, P. (2002). Coal mining and the environment. *Energy Research Institute, University of Cape Town*.

Lori, M. (2009). Data security in the world of cloud computing. *Co-published by the IEEE Computer and Reliability Societies*, pages 61–64.

Meek, A. (2013). A systems engineering approach to improve the measurement and verification process of energy services companies. Master's thesis, Faculty of Engineering, North-West University.

Melnikov, A. and Zeilenga, K. (2006). Simple Authentication and Security Layer (SASL). Technical report, RFC 4422, June.

M.I. Howells (2006). The targeting of industrial energy audits for DSM planning. *Journal of Energy in Southern Africa*.

Microsoft (2014). SQL Server 2014 editions,
http://msdn.microsoft.com/en-us/library/cc645993.aspx.

MySQL (2014). MySQL editions,
http://www.mysql.com/products.

National Energy Regulator of South Africa (2010). NERSA Consultation Paper, Revision of Regulatory Rules for Energy Efficiency and Demand Side Management (EEDSM) including Standard Offer Programme (SOP). Kulawula House, 526 Madiba Street, Arcadia, Pretoria, 0001.

Nikiforakis, N., Balduzzi, M., Van Acker, S., Joosen, W., and Balzarotti, D. (2011). Exposing the Lack of Privacy in File Hosting Services. *Proceedings of the 4th USENIX conference on Large-scale exploits and emergent threats (LEET'11).*, pages 1–8.

Oracle (2014). Oracle database editions,
http://www.oracle.com/us/products/database/enterprise-edition/comparisons/index.html.

Palensky, P. (2003). The jevis system: An advanced database for energy-related services. *Institute of Computer Technology, Vienna University of Technology*, pages 1–5.

Park, J. and Devarajan, G. (2007). Fine-grained and scalable message protection in sensitive organizations. *Journal of Software*, pages 64–75.

Patel, K., Utareja, S., and Gupta, H. (2013). Information hiding using least significant bit steganography and blowfish algorithm. *International Journal of Computer Applications*, pages 24–28.

PostgreSQL (2014). PostgreSQL, http://www.postgresql.org.

Powertech (2014). Powertech System Integrators, http://www.ptsi.co.za.

Raja, S. (2012). Securing risks of electronic mail based on the type of organization. In *International JOurnal of Innovation, Management and Technology*, pages 246–249.

Ramgovind, S., Eloff, M., and Smith, E. (2010). The management of security in cloud computing. In *Information Security for South Africa (ISSA), 2010*, pages 1–7. IEEE.

Republic of South Africa (2006). *Regulation Gazette No. 28992*, volume 493. Government Printing Works, Pretoria.

Republic of South Africa (2013a). *Regulation Gazette No. 35069*, volume 582. Government Printing Works, Pretoria.

Republic of South Africa (2013b). *Regulation Gazette No. 37136*, volume 582. Government Printing Works, Pretoria.

SABS Standards Division (2011). SANS 50010, Measurement And Verification Of Energy Savings. 1 Dr Lategan Road, Groenkloof, Pretoria, 0001.

SANEDI (2011). Regulations On The Allowance For Energy Efficiency Savings. 150 Linden Street, Strathavon, Sandton, 2146.

Sanjuàs-Cuxart, J., Barlet-Ros, P., and Solé-Pareta, J. (2012). Measurement based analysis of one-click file hosting services. *Journal of Network and Systems Management*, pages 276–301.

Schneider Electric – Earth Rangers (2011). Energy Management Information System (EMIS) Case Study, Earth Rangers. http://www2.schneider-electric.com/socuments/solutions/struxureware/Earth_Rangers_Case_Study.pdf.

Schneider Electric – Fastweb (2010). Energy Management Information System (EMIS) Case Study, Fastweb. http://www2.schneider-electric.com/documents/solutions/struxureware/Fastweb_Case_Study.pdf.

Schneider Electric – Sanofi Pasteur (2012). Energy Management Information System (EMIS) Case Study, Sanofi Pasteur. http://www2.schneider-electric.com/documents/solutions/struxureware/Sanofi_Case_Study.pdf.

Shao, F., Chang, Z., and Zhang, Y. (2010). AES encryption algorithm based on the high performance computing of gpu. In *Communication Software and Networks, 2010. ICCSN'10. Second International Conference on*, pages 588–590. IEEE.

Simitsis, A., Vassiliadis, P., and Sellis, T. (2005). Optimizing ETL processes in data warehouses. In *Data Engineering, 2005. ICDE 2005. Proceedings. 21st International Conference on*, pages 564–575. IEEE.

Singh, S. and Maini, R. (2011). Comparison of data encryption algorithms. *International Journal of Computer Science and Communication*, pages 125–127.

Spalding-Flecher, R. (2006). Demand-side energy efficiency promotion under the Clean Development Mechanism: lessons learned and future prospects.

SQLite (2014). SQLite,
http://www.sqlite.org/index.

Statistics South Africa (2005). Consumer Price Index (CPI). Private Bag X44, Pretoria, 0001.

Stephen, R. (2009). *Demand Side Management in South Africa*, volume 2. Crown Publications.

U.S. Energy Information Administration (IEA) (2014). Overview of South Africa. http://www.eia.gov.

Van der Zee, L. (2013). Modelling of electricity cost risks and opportunities in the gold mining industry. Master's thesis, Faculty of Engineering, North-West University.

Vassiliadis, P., Simitsis, A., Georgantas, P., Terrovitis, M., and Skiadopoulos, S. (2005). A generic and customizable framework for the design of ETL scenarios. *Information Systems*, pages 492–525.

Vassiliadis, P., Simitsis, A., and Skiadopoulos, S. (2002). Conceptual modeling for ETL processes. In *Proceedings of the 5th ACM international workshop on Data Warehousing and OLAP*, pages 14–21. ACM.

Velázquez, D., González-Falcón, R., Pérez-Lombard, L., Marina Gallego, L., Monedero, I., and Biscarri, F. (2013). Development of an energy management system for a naphtha reforming plant: A data mining approach. *Energy Conversion and Management*, pages 217–225.