A mobile application approach for improving data consolidation in industry

JL Skinner

orcid.org/ 0000-0001-5176-755X

Dissertation accepted in fulfilment of the requirements for the degree *Master of Engineering in Computer and Electronic Engineering* at the North-West University

Supervisor: Dr JF van Rensburg

Graduation: November 2023

ACKNOWLEDGEMENTS

Thank you for your unending love and support, Anel Skinner, my wife. Every day I find you to be an inspiration. Nothing would have been possible without your steadfast support.

Thank you for your love and support, Mom and Dad. I appreciate your never losing faith in me and helping me develop self-confidence. Your affection and assistance are much valued.

I would like to express my gratitude to Dr JF Van Rensburg, Dr J Prinsloo, and Dr J van Laar for their invaluable time and contributions to the writing of this dissertation.

Thank you for letting me complete my master's at CRCED Pretoria, Prof EH Mathews.

I want to thank all of my family and friends for their help and patience.

ABSTRACT

Reliable data collection is crucial in South Africa's industrial sector due to the escalating operational costs. Effective data utilisation allows companies to optimise their systems, enhance resource efficiency, and expedite decision-making processes. Leveraging data-driven insights empowers companies to identify opportunities for improved resource allocation, leading to cost savings and enhanced performance. To address the need for robust data collection in challenging environments, a curated application was developed and integrated with an existing online system. This integration ensures real-time data synchronisation and efficient access, facilitating seamless data flow and system performance. Through analysing existing data collection applications, deficiencies were identified, and shortcomings were addressed, particularly in the storage and processing methods employed.

The developed cross-platform application utilises onboard components across different devices and platforms. The artefact seamlessly integrates with existing systems to store data on a server. The application dynamically generates collection pages based on active generation, allowing for diverse data collection configurations without altering the application structure. Moreover, it generates adaptable storage structures for various data layouts, ensuring flexibility and expandability.

The system was then assessed using a case study conducted at South African mines, specifically focusing on collecting diesel log readings from mining equipment and vehicles. The system was configured to meet the specific needs of each mine, including the functionality to scan barcodes attached to each asset. The system successfully collected the data and saved user, location and date information ensuring traceability. The collected data was utilised for various purposes, including diesel rebate processing.

The diesel rebate claims, created using the data collected with the system, were accepted by the South African Revenue Service (SARS). In addition, the inclusion of pictures and user information in the logs further enhanced the reliability of the data. The diesel rebates resulted in substantial cost savings for the mines, totalling an average of R1.5 million per month (to date).

This dissertation addresses the critical requirement for reliable data collection in South Africa's industrial sector. The developed cross-platform application demonstrates its efficacy in traceable data collection, accommodating diverse configurations and seamless integration with existing systems. The case study at South African mines showcases the system's ability to efficiently collect and process data, resulting in significant cost savings through successful diesel rebate claims. Overall, this research contributes to advancing data collection methods, benefiting industrial operations in South Africa.

Title: A mobile application approach for improving data consolidation in industry

Author: Jeffrey Skinner

Promoter: Dr Johann Van Rensburg

Co-Promoter: Dr J Prinsloo

Keywords: Mobile Data Collection, Data Consolidation, Mine Operations Optimisation, Offline Data Syncing

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

1 INTRODUCTION

1.1 Background

1.1.1 Industrial and economic environment

Since this thesis will be focusing on the data collection in the South African industrial sector, this chapter will start by discussing some of the aspects of the sector. The South African industrial sector is currently facing challenges that are impacting its competitiveness in the global markets [1, 2]. The Department of Communication [1] and Lane *et al*. [2] have highlighted various issues including declining ore grades, decreasing commodity values, lower productivity, labour disruptions, and rising operational expenses. These issues have also been acknowledged by Marais *et al*., and Groenewald *et al*. [3, 4], who emphasised the pressure on the South African industrial sector to remain viable, amidst labour disruptions, commodity value decreases and increasing operational expenses. Pressure from the factors stated ultimately leading to economic degradation [3, 4].

In addition to these challenges, the South African industrial sector is concerned with increasing electricity prices. Before 2008, electricity prices in South Africa were low and decreasing. However, a severe power supply shortage in 2008 prompted Eskom, South Africa's electricity utility, to implement a new program. This program aimed to increase power generation capacity [3]. As a result of the new program electricity prices were increased, by 2016, electricity prices had surged by 290% more than the equivalent increase experienced in 2000 [5]. Figure 1, based on data from Eskom [6], is supported by the study conducted by Mokoena *et al*. [7], which depicts the electricity consumption compared to inflation. This graph was generated with data illustrating the rise in electricity cost as well as the increase adjusted for inflation. Even when adjusted for inflation, the graph demonstrates a noticeable increase. Moreover, the escalated electricity tariffs in South Africa have adversely affected the capital expenditure and profitability of industries reliant on electricity [8].

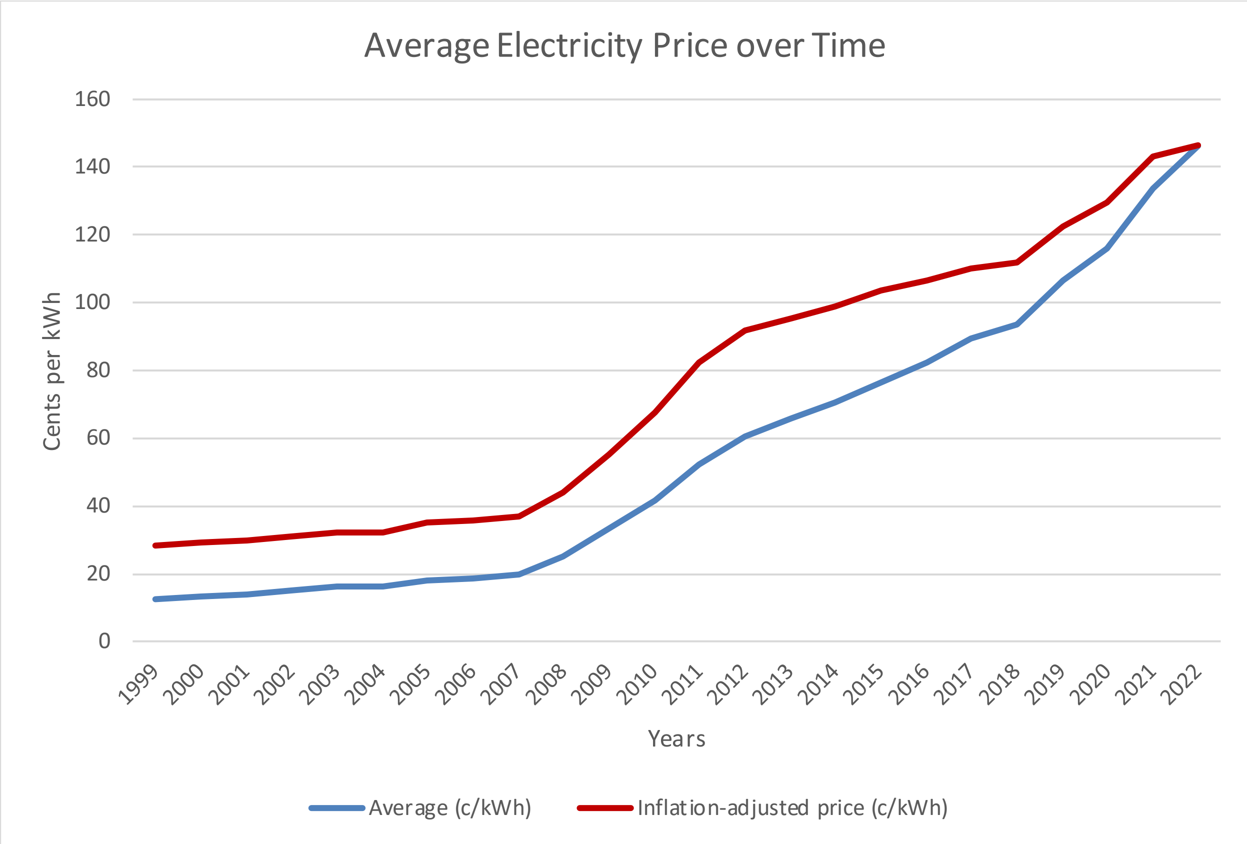


Figure 1: Average electricity price in South Africa since 1999 [9, 10]

Another study focusing on the South African mining sector revealed a significant increase in input costs. The annual average input cost escalation in 2021 stood at 8.58% year-to-date, as reported by Mining Journal and Minerals council South Africa [9, 10]. This is compared with:

- 6.61% in 2020
- 8.26% in 2019
- 8.67% in 2018

With consistent increases in input costs, as indicated by these studies, and decreasing commodity values, as stated in [1, 2], South African industries must find a balance to mitigate the losses resulting from these challenges.

Optimising existing systems is one crucial approach to offsetting the losses incurred from these challenges. By optimising systems, operational costs can be reduced through resource efficiency and decreased production time [11, 12]. Data about the system plays a pivotal role in achieving system optimisation. In the following section, we delve deeper into the data requirements of the industrial sector.

1.1.2 Industrial data needs

To optimise an existing system, data about the current system is needed [13]. The work done by Goosen *et al*. [13] focussed on creating a system to quantify stored data in the South African industrial sector. Supported by the contributions of Herman *et al*. [14], a conclusion was reached that a data-capturing mechanism can be of use to industries only when data is stored, processed, and analysed. Using data in this way can add and already has added value to many South African industries [13].

Furthermore, the utilisation of data brings numerous benefits to industries, including cost reduction, production enhancement, and system optimisation. Data analysis allows companies to identify inefficiencies, streamline processes, and allocate resources more effectively, leading to cost savings. Data-driven insights enable companies to make informed decisions, improve operational efficiency, and increase productivity. By analysing data, companies can identify bottlenecks, optimise workflows, and implement targeted improvements, resulting in increased production output. Overall, the strategic use of data has positively impacted South African industries by driving cost savings, improving productivity, and optimising system performance [15].

The South African industrial sector already collects vast amounts of data [4]. Due to a lack of resources such as an internet connection and the inconsistency of the South African electricity supply, the data is mainly stored as paper-based logs [16]. Data storage methods and their strengths and weaknesses will be discussed in section 1.1.4.

The data collected by the South African industrial sector is increasing [17, 18]. This increase is motivated by two factors. Firstly, data can be used to start projects that need data to prove the validity or reliability of statements or ideas. In addition, new projects could have a monetary value. Secondly, previously unused data are now being reported because of legislative changes in the sector [17, 18].

The amount of data collected and the method by which the data is collected can decrease the data quality and usability [14]. In Figure 2, the global trend for data collection can be seen. Reinsel *et al*. [19] predicted that 150 billion devices will be connected by 2025, and 30% will be real-time data collection devices. The importance of data quality will be discussed in section 1.1.3. The data gathered in many South African industries are not always traceable, and legislative changes require that data be traceable. With these changes, South African industries might be compelled to look at their data collection methods to stay compliant.

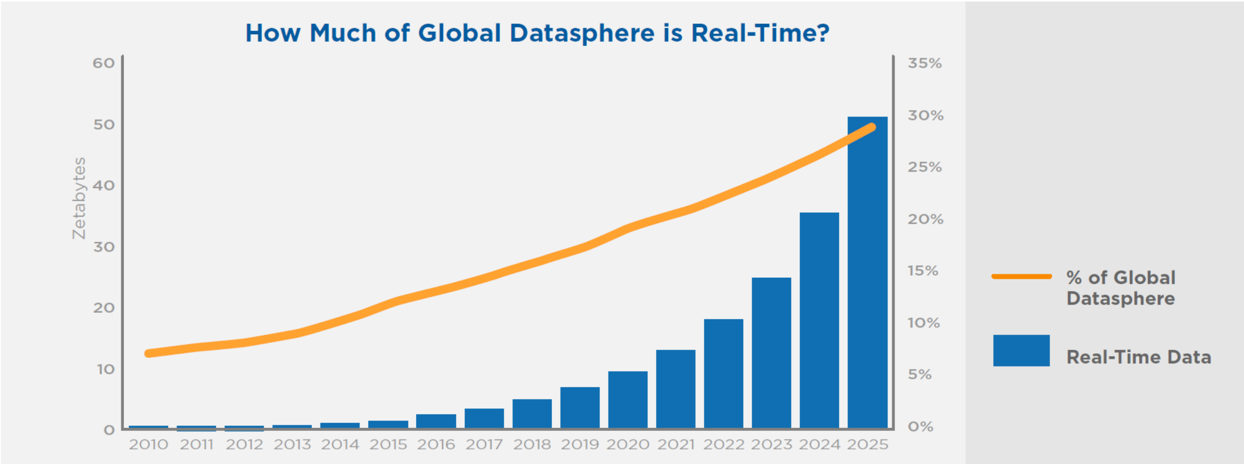


Figure 2: Increasing real-time data needs since 2010 [20]

Improvements in productivity, efficiency, quality, and the integration of the construction process are the measures for the transformation's positive effects on the industrial sector. Moreover, they can remove obstacles that could prevent the building industry from digitising [20].

Two topics that require more investigation were found by investigating the South African industrial sector's data needs. Firstly, the data's storage method significantly influences the data's usability. Secondly, the data quality must be high enough to help ensure that the data can optimise systems accurately. The following section will investigate data quality and traceability.

1.1.3 Data quality and traceability in industrial systems

Data quality is crucial for businesses as it directly affects their ability to make informed decisions and reach cost saving targets. This section explores the factors that determine data quality, the costs associated with poor data quality, the benefits of investing in data quality management, and the potential for Industry 4.0 and data consolidation to enhance business-supplier relationships (BSR).

Data quality can be described as the ability to meet the industry's requirements. Some of the factors that determine the ability of the data to meet industry requirements are summarised in Table 1 [21, 22]:

Table 1: Data quality properties and their descriptions

Name of property	Description
Accuracy	The degree to which a value or a standard conforms to a measurement, a specification or a result of a calculation
Accessibility	The ability of the data to be reached or opened
Availability	The quality of being able to be used or obtained
Believability	The quality of being convincing or realistic
Completeness	The state or condition of having all the necessary or appropriate parts
Compliance	The ease with which the data conforms to given instructions
Consistency	The quality of acting or being done in the same way
Integrity	Strong moral principles or honesty
Objectivity	Lack of favouritism toward one side or another
Relevance	The quality or state of being intricately connected or appropriate
Reliability	The degree the data can be trusted to implement a result of a calculation, a measurement, or a specification
Timeliness	Occurs at a favourable or advantageous time
Validity	Acceptable in an official or legal manner

Poor-quality data can be very costly, and several companies and projects have lost money because of low-quality data. Poor data quality can have a side effect that leads to money loss with factors like [23, 24]:

- Possible increase in costs associated with data management
- Possible increase in costs for data retrieval
- Costly delays in the delivery of goods and services
- Increased upkeep expenses
- Increased legal expenses
- Increased labour expenses
- Missed opportunity and lost revenue expenses
- Process failure expenses
- Operational expenses
- Time spent on useless information expenses

In his dissertation, C.J. Bester [25] highlights the financial challenges faced by Eskom and the importance of data quality in making informed decisions and achieving cost savings. The researcher surveyed 303 Eskom technicians and analysed historical data on customer calls related to electricity supply problems (ESP) to determine the impact of data quality on cost savings. The results revealed potential monetary savings of 17.18% by improving transactional data quality on ESP customer calls. The study suggests that if the costs of improving data quality are lower than the potential savings, Eskom's financial position could be improved. Additionally, the analysis indicates that improving customer call transactional data quality at the source could result in savings of up to 47.7% for ESP-related transactions. The study recommends that Eskom continues to prioritise high-value transactional data quality to achieve significant cost avoidance and savings potential [25]. The other point is that data standards are ever-increasing, and the industry expects better-quality data constantly [21, 22].

The success and operations of an organisation can suffer from poor data quality, due to factors such as data inconsistency, data duplication, data integration complexities, lack of data governance, and insufficient data validation processes. Organisations navigate through these obstacles to ensure that the data they collect is accurate, reliable, and consistent. Strategies for overcoming these challenges could involve implementing robust data quality management frameworks, establishing clear data governance policies, investing in data cleansing and validation tools, conducting regular data audits, and fostering a data-driven culture within the organisation. By addressing these challenges head-on, companies can enhance their data quality management practices and derive greater value from their data assets.

Data quality, however, is a two-way street. Managing and making investments in data quality has been identified by businesses as having numerous real and tangible benefits. These advantages comprise the following [13, 23]:

- Establishing a single version of the truth
- Boosting analytical systems confidence
- Increasing customer satisfaction
- Decreasing cost and increasing income
- Decreasing time spent reconciling data

According to Hossan *et al*. [16], businesses that treat their data as a strategic resource and make quality improvements gain an edge over rivals that do not. Thus, businesses should improve their data quality management efforts to benefit from high-quality data.

Tomlinson *et al*. [26] noted that immediate data access significantly improves quality. Furthermore, data collection can be guided by implementing advanced skip patterns based on entered values.

A qualitative content analysis was performed by Viele *et al*. [27] to find similar patterns, themes, and categories based on empirical data collected from 45 expert interviews with executives from German and Austrian industrial enterprises. Based on these interviews, Viele *et al*. [27] surmised that future transactions will be based primarily on digitised, automated processes, and various value-creation methods will be shifted to platforms. BSR will intensify by concentrating on important strategic suppliers, and businesses can streamline their supplier network. This helps companies to leverage the full potential of Industry 4.0 for BSRs to create and secure long-term and sustainable competitive advantages, giving potential for Industry 4.0 and data consolidation to develop better relationships and projects [27].

In this section, an investigation into the data quality and usability of the data was done. It is determined that data quality is a critical aspect of effective decision-making and cost savings for businesses. Overcoming challenges in data quality management and harnessing the potential of Industry 4.0 and data consolidation can further drive improvements in business processes, relationships, and competitiveness. To ensure that the data quality stays high, error-handling techniques and redundant data (data in different formats) to decrease user input errors will be required. This will be discussed later in this chapter. This section also found that the method by which data is collected and stored significantly influences the data quality. The data consolidation methods used in the South African industrial environment will be discussed in the next section.

1.1.4 Industrial information systems

Data storage plays a crucial role in the effective management and utilisation of information collected in South African industrial systems. Choosing the appropriate storage method is essential for ensuring data accessibility, usability, and long-term preservation. This section examines two primary storage methods: physical paper-based storage and digital repositories stored on a server. By exploring the advantages and disadvantages of each approach, we can better understand the challenges faced by industries in storing and managing their data effectively.

Data can be stored in a physical format, such as paper. In such cases, digital data that requires physical storage will need to be printed. According to a report by Eskom [28], physical data storage becomes burdened with administrative tasks such as processing and analysing physical data. Data stored in this way is hard to filter and find easily. This makes the data less usable and sometimes costly to obtain. This information also needs to be physically located and retrieved to be used, and this process can be very time-consuming [29, 30].

South Africa's Department of Home Affairs has a lot of delays contributed to paper-based logs [31]. According to the Department of Home Affairs, it holds more than 350 million paper copies of civic records from 1895 that include information on births, marriages, deaths, and amendments. According to the article [31], searching through these documents to find requested documents causes significant delays for the Department and the South Africans that require the requested records. The Department of Home Affairs has declared that it will recruit 10,000 unemployed young graduates starting November 2022 to digitise these paper records because this has become such a problem. Starting salaries for these workers range from R5,000 for data loggers to R14,250 for manager-level positions. The project is expected to last at least six months and might cost the department at least R500 000 each month. According to the government, the value of having digital copies will optimise the system and reduce the current backlog [31].

Along with the administrative problem, another problem was pointed out by Schobel *et al.* [29]. Practically speaking, paper-based surveys need physical storage space, which can complicate the logistics of storing and retrieving the data [29]. Instream [30] has shown that paper-based data collection systems are expensive to maintain on a wide scale and that long turnaround times and dubious accuracy are once again underlined. In addition, Kinkade *et al*. [32] bring up the issue that manual data collection techniques are vulnerable to fraud. Storing data in physical form might seem the cheapest of the options at first, but this does not seem to be the case given the administrative and logistical challenges.

The second option is to digitise storage, using online storage methods to store the collected information as a possibility. This creates a centralised location where the data can always be accessed from anywhere with a connected communication medium. In addition to data analysis functionalities, a central repository is required to enable other advanced functionality, including system integration and distributed data management [26]. Using online storage also makes filtering and finding required data more accessible. Since no physical data needs to be moved around, the space needed to store digital information is less at larger scales. A survey was conducted, in which participants were asked about the most significant issues with a paper-based process in their organisation. The results of this study can be seen in Figure 3. Most participants said their most prominent issue was the time spent re-keying data or searching for copies. Online data storage removes many of the administrative and logistical challenges faced by physically storing data [33].

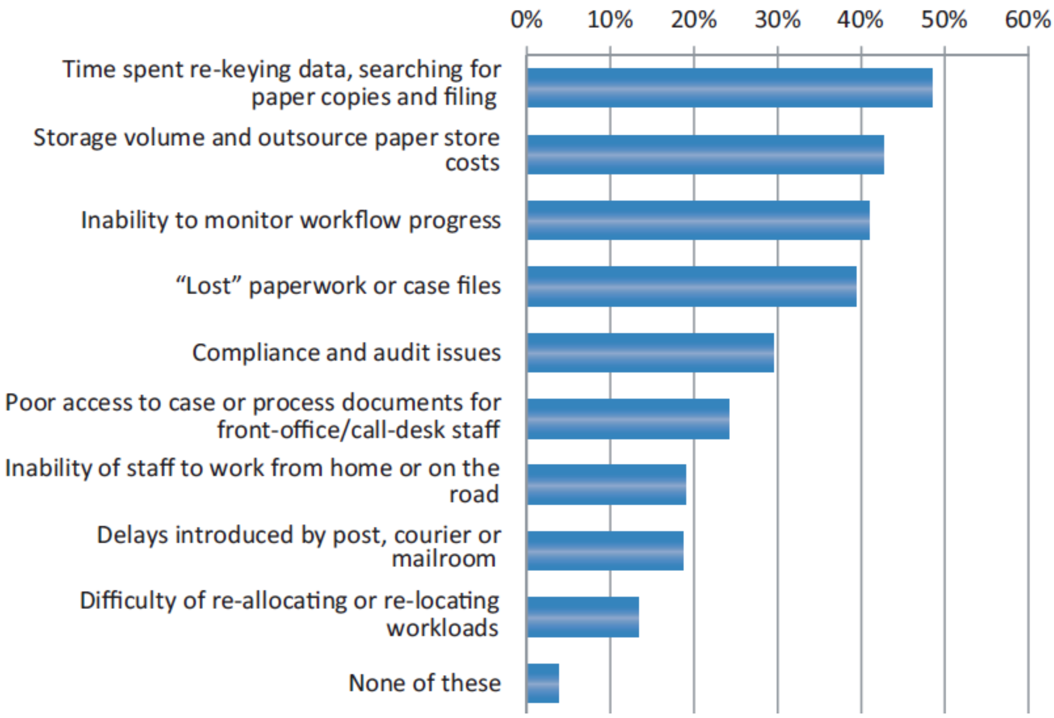


Figure 3: Biggest issues caused by paper-based processes [33]

More *et al*. [34] did a study on the application of digital technologies on mine sites in South Africa. This study discussed how smart mine water management requires the continuous advancement of Industrial-Internet of Things (IIoT) technologies to optimise the mining process and address environmental issues. To achieve full-scale production and reduce costs, continuous data collection, automation, machine-to-machine communication, and increased data analysis are essential [34]. As outdated mining methods are proving to be costly, there is a need for "intelligent" mining through interconnected data communication and advanced analytics. The use of digital data management along with sensor technology, 3D modelling, and robotics can aid in detecting harmful metals in underlying rocks and improve mine water management. In this context, data storage plays a crucial role, and mines are increasingly adopting comprehensive solutions that leverage highly advanced analytical techniques. Data lakes are considered more suitable than data warehouses for predicting future mine water scenarios. More *et al*. [34] suggest that future research should focus on Big Data analytics across mining processes, emphasising dynamic modelling approaches for improved mine water balance. These advancements in data storage and analysis will enhance mine water management and make the transition to smart mining more attractive for future operations [34].

Storing information online has a few challenges. If the data is collected on paper or other non-digital mediums, the data will have to be digitised. The primary challenge which the South African Department of Home Affairs faces, mentioned in Motsoaledi *et al*. [31], is that digitising collected data can also become costly. Another challenge when storing information digitally is the connection to the online server. A communication medium is required to get the data to the cloud-based storage. The data will need to be uploaded using an existing Application Programming Interface (API), so data collected where the internet or other communication methods are not viable must be synchronised later [26, 35, 36]. With the constraints mentioned for the South African industrial sector, if the cloud-based storage option is chosen, the data should be collectable offline and synchronised later in addition to being uploaded in real-time as a safety precaution.

The challenges of industrial information systems and data storage methods are closely related to the concept of Big Data. In the first part of this section, we discussed the challenges associated with physical paper-based storage and the limitations it poses in terms of data accessibility, usability, and long-term preservation. Similarly, digital repositories stored on a server were introduced as an alternative storage method, offering advantages such as improved data filtering, accessibility, and reduced administrative and logistical challenges. However, to fully understand the significance of Big Data, it is crucial to recognise its role in addressing these challenges. Big Data analytics enables industries to process and analyse substantial amounts of data, characterised by its value, variety, volume, velocity, veracity, and complexity [38].

Big Data or Big Data analytics is a new trend in industries. It refers to technology processing substantial amounts of data with six main characteristics namely: Value, Variety, Volume, Velocity, Veracity, and Complexity [37, 38]. In addition, Big Data includes unstructured (unorganised and text-heavy) and multi-structured data (human-machine interaction leading to various data formats) [39]. With the advent of sensors for data acquisition in drilling, exploration, and production operations as it pertains to various mining sectors, industries have become data-intensive. Although enterprises have recently shown increased interest in using Big Data analytics, there are still challenges, mainly due to the lack of corporate support and awareness of Big Data in industries. In addition, the data's quality and understanding of the problem's complexity are among the challenges in applying Big Data [40].

- **Value** is an essential feature of Big Data. Investments in Big Data infrastructures have a significant return on their value. Big Data analyses enormous data sets to find underlying patterns and assist engineers in foreseeing potential issues. Understanding the future performance of operational assets and spotting problems can provide the organisation with a competitive edge and increase the value of the company [40].
- **Variety** refers to the several types of data generated, stored, and analysed. Due to the many types of data-collecting devices and sensors, the collected data may come in a variety of sizes and formats. The generated data can be in text, image, audio, or video

format. It is technically possible to divide the classification into organised, semi-structured, and unstructured data [41]. It is reported that 90% of generated data is generally unstructured [38].

- **Velocity** as a feature of Big Data relates to the rate at which data is transferred and processed. Also, it alludes to how quickly data is produced. The problem with the speed component is the small number of processing units compared to the available volume of data. Yet, recent times have seen a remarkable increase in data generating pace, with 5 exabytes produced in just two days. This represents the entirety of human history's data up to 2003 [41].
- **Veracity** is a term used to describe the standard and value of the data that is available for analysis and decision-making. The contrast between clean and unclean data is also included. This is particularly crucial because messy data can dramatically slow down and impair data processing. As well as being professionally and effectively processed and filtered for data analysis, the generated data must be accurate. Otherwise, the findings will not be trustworthy. Another challenge is that the data is collected by conventional manual data collection by human personnel and could therefore contain inaccuracies or mistakes [40]
- **Volume** refers to the amount of data or information. This data can come from any sensor or data acquisition device. Handling these enormous volumes of data is challenging due to problems in storage, sustainability, and analysis [39]. Many organisations have vast amounts of data in their archives but cannot process it. The main application of Big Data is to provide processing and analysis tools for growing data volumes [38].
- The **complexity** of the problem for which the data collection is performed is significant [42]. Dealing with large data sets from a complex data processing problem is challenging, and it can be difficult to find the underlying trend. For these problems, Big Data tools can be beneficial [40].

Figure 4 summarises the characteristics of Big Data mentioned above.

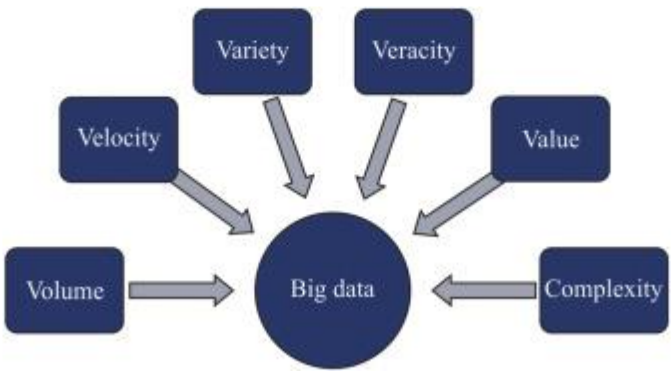


Figure 4: Big Data characteristics [40]

All industries, including automotive, manufacturing, oil and gas, aerospace, healthcare, retail, and mining, to mention a few, have developed various projects to introduce Big Data technologies into their operations. Yet, the adoption of Big Data technology has been slower in the extractive industries. The Google search engine tries to count the number of articles that reference "Big Data" and a specific industry: Table 2 can be used to confirm this. To separate the mining sector from "data mining," the term "mining industry" is used. The quantity of Google search results reveals how well-known and well-liked Big Data technologies are in the relevant business [43].

Table 2: String for Google Search by Industry [43]

Industry	*Search String*
Manufacturing	"big data" AND "manufacturing"
Automotive	"big data" AND "automotive"
Aviation	"big data" AND "aviation"
Healthcare	"big data" AND "healthcare"
Retail	"big data" AND "retail"
Oil & Gas	"big data" AND "oil and gas"
Mining	"big data" AND "mining"

The results of this Google search test are shown in Figure 5. The industries that are most interested in Big Data technologies are the industrial and healthcare sectors. Interest in Big Data technologies is seen in current trends in personalised healthcare and autonomous vehicles. The mining, aerospace, and oil and gas sectors are less interested in implementing Big Data [43].

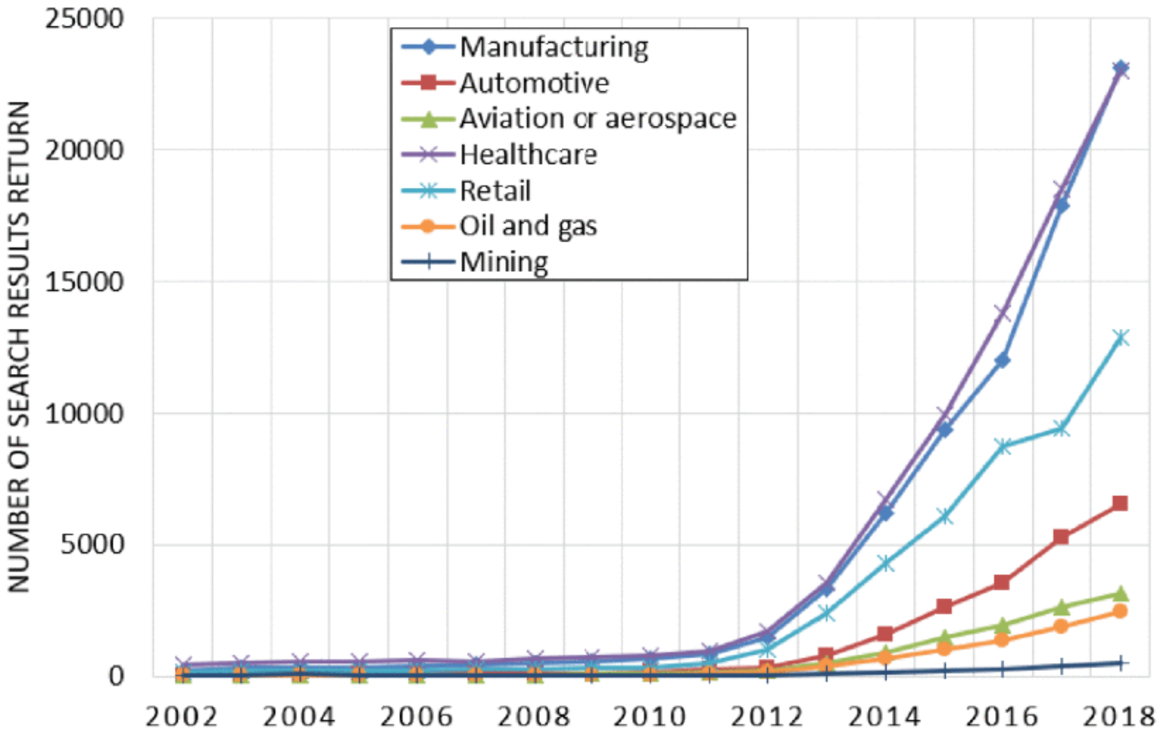


Figure 5: Google search results for "big data" in each industry [43]

Current Big Data technology implementations have increased operational effectiveness and maximised asset value. The capacity of computers has expanded thanks to advancements in data processing, storage, and cloud technologies, creating new possibilities for real-time monitoring applications like smart refineries. The emergence of digital twins, which is a virtual representations designed to simulate a physical object, wearable computers, artificial intelligence, blockchain, and robotics are further examples of digitisation [43].

In this section, it became clear that the method of collecting data makes a difference in storing it, and Big Data and how it influences industries were explained. Therefore, data consolidation systems are investigated in the next section to ensure that the data collection and storage methods could work well together.

1.1.5 Industrial data consolidation systems

In the South African industrial sector, where large amounts of data are generated, consolidated data systems are vital for optimising operations and achieving efficiency. These systems enable the integration and organisation of diverse data from different sources, facilitating comprehensive analysis and generating meaningful insights. With data consolidation, organisations can overcome the challenges associated with scattered and fragmented data, ensuring data traceability, accuracy, and reliability [44, 45]. Various methods are used to collect data in the South African industrial environment. The three main types of systems investigated in this section are automatic data acquisition systems, manual non-digital data collection and manual digital data collection.

The first option is an automatic data acquisition system known as supervisory control and data acquisition (SCADA). A SCADA can be set up to monitor a set point and consolidate the information gathered [44, 45].

Argawal *et al.* [46] found the following advantages of SCADA systems:

- SCADA systems are robust because SCADA systems are used in critical industrial process environments where performance and reliability are crucial
- Large amounts of process data can be recorded and stored by SCADA systems
- Simple interfaces offer great functionality due to strong development history
- Employees have access to process data in real-time
- A limited number of specific developments must be carried out by the end-user

However, Krasner *et al*. [47] believe that although integrating SCADA systems may be advantageous, doing so still necessitates a large financial investment and takes time to do so effectively. Since SCADA systems are set up at a set point, the system will also be unable to collect event-based data beyond the point of the initial setup. Collecting data based solely on human interaction techniques will not apply to this method.

The second method to consider is manual non-digital data collection. This data collection method involves sending personnel to the data collection event and using paper logs to collect data. Manual data collection is typically performed using a pen-and-paper approach. Lwin *et al*. [35] indicates that traditional data collection is an uphill task, highlighting specific challenges that contribute to its limitations. One key issue with manual non-digital data collection is the lack of traceability regarding the individuals who collect the data. Since the data is not digitally recorded, it becomes difficult to attribute the data to specific personnel, compromising accountability and quality control [35].

Moreover, the data collected through this method remains in a non-digital format, making it less compatible with modern data analysis and optimisation techniques. To make the data collected in this manner more usable and effective for tasks such as system optimisation and tracking, it typically needs to be digitised. However, the process of digitising data is time-consuming and prone to errors [35, 48]. Introducing digitisation to convert the data into a digital format adds another layer of complexity, potentially introducing inaccuracies and reducing the reliability of the data [33, 46].

As discussed in previous sections, relying on paper-based logs for data storage and retrieval poses numerous administrative and logistical challenges. The need to physically handle and store paper-based logs introduces additional time-consuming tasks and increases the risk of misplacement or damage, further impeding the efficiency and reliability of the data collection process. Therefore, it is evident that traditional manual non-digital data collection faces various challenges that hinder its efficiency and reliability. These challenges include the lack of traceability, the need for digitisation, and the administrative and logistical complexities associated with paper-based logs [33, 46].

The last method of data collection to consider is manual digital data collection. This data collection method uses personnel and sends them to collect data with a digital device such as a smartphone. Smartphones have various input forms and tools that help facilitate data collection such as cameras, Global Positioning System (GPS), and internet connection. Users can use these multiple input methods to make data more reliable by including supporting information with the collected data [36].

This data can also be consolidated using the smartphone's communication systems. With the use of mobile applications and cloud-based platforms, workers can input and synchronise data in real time from their smartphones, ensuring that the data is collected and consolidated efficiently. This allows for seamless communication and collaboration among team members, regardless of their physical location. The ability to consolidate data through smartphone communication systems provides convenience and accessibility, enabling organisations to make informed decisions based on up-to-date information.

It is important to note that manual digital data collection may not be as effective in industries where static data is primarily required. In such cases, alternative methods like SCADA systems, which are designed for monitoring and recording large amounts of process data, might be more suitable. However, for industries that have event-based data collection needs, manual digital data collection using smartphones proves to be a valuable and efficient approach. This method requires that a mobile application be written for the device that can collect this data and that personnel are trained in using the application [36].

The operation sequence of a typical mobile data collection system with consolidated data was described by Park [36]:

1. Design forms for data collection
2. Give out forms to field devices
3. Collect data with mobile devices
4. Submit data to a centralised database
5. Process data as it is received
6. Provide results and reports

The data collection methods and data storage methods have been discussed in this section. The next step would be to design a system that would allow for the most optimum use of this information, depending on the industry. To ensure the design will be able to meet this requirement the next section will look at physical and digital design considerations used in the industry.

1.1.6 Design best practices in the industry

In this section, two design elements will be investigated, namely the front-end and the back-end design. The back-end design will be the storage and collection methods and how they best work together. The front-end design would be what the user physically sees and how the user can best collect the data.

Back-end system design

A trustworthy system must have both reliability and security, yet it is challenging to build both features into systems. Although the criteria for security and reliability have many similar aspects, they also call for separate development considerations. Furthermore, it is simple to ignore the complex interactions between safety and reliability that can produce unanticipated outcomes [49].

Although security and reliability both focus on the confidentiality, integrity, and availability of systems, security and reliability have differing perspectives on these qualities. The existence or absence of a hostile enemy is the primary distinction between the two points of view [49].

In terms of reliability, several factors contribute to maintaining a robust and dependable system. For instance, mechanisms such as load balancing and load shedding are implemented to distribute the workload across multiple servers and prioritise critical tasks. Load balancing ensures that requests are evenly distributed, preventing any single server from being overwhelmed and reducing the risk of service disruptions. On the other hand, load shedding involves prioritising certain tasks or services over others during high-demand periods to ensure optimal performance and resource allocation [49, 50]. Additionally, communication within the system relies on components such as routers, switches, and protocols like TCP/IP, which facilitate the seamless transmission of data between different nodes. These components play a crucial role in forwarding requests, ensuring timely delivery, and maintaining the availability of the system. By incorporating these mechanisms and configurations, the system can effectively handle high traffic volumes and mitigate the risks of performance degradation or downtime. Concrete examples of load balancing algorithms, such as round-robin or weighted round-robin, and techniques for load shedding, such as prioritising real-time data over non-critical data, can further enhance the reliability of the system. It is important to consider these specific mechanisms and components when designing a reliable and secure system to meet the challenges of data collection and communication in various industries [50].

Reliability and security are interconnected and heavily reliant on the availability of the system. However, ensuring system availability goes beyond mere uptime and encompasses protection against potential vulnerabilities that could be exploited by malicious attackers. To enhance security, various measures can be implemented. Anomaly detection systems play a crucial role in identifying abnormal patterns within the system, enabling prompt detection and response to potential security breaches [51]. These systems utilise advanced algorithms and machine learning techniques to analyse network traffic, user behaviour, and system logs, identifying any suspicious activities that may indicate a security threat. Audit logging is another essential component of a secure system, as it enables the tracking and recording of all user actions and system activities [51]. By maintaining detailed logs, organisations can trace security incidents, investigate potential breaches, and ensure accountability. Security monitoring tools provide real-time visibility into the system, monitoring network traffic, system logs, and user activities for any signs of unauthorised access, malware infections, or suspicious behaviour [51]. These tools provide alerts and notifications, allowing security teams to respond swiftly to potential threats and take necessary actions to safeguard the system. By implementing robust anomaly detection, audit logging, and security monitoring tools, organisations can proactively protect against attacks, maintain system availability, and safeguard authorised user operations [50].

One example of such an attack is a distributed denial-of-service (DDoS) assault, where many devices worldwide are commanded to flood a target with excessive traffic, overwhelming its resources and causing service disruption. By considering the challenges related to confidentiality, integrity, and availability, the proposed solution for mobile data collection and log keeping in mines must incorporate robust security measures. This includes encryption, access controls, and network resilience to mitigate the risks of unauthorised access, data breaches, and potential disruptions. Safeguarding user data and ensuring the availability and integrity of the system are critical factors in developing a reliable and secure mobile data collection solution for mining operations [50].

A malicious attack cannot be distinguished from a design flaw or a legitimate increase in traffic from the victim's perspective. Like a classic application-level DDoS assault, a significant message or other events that prompt millions of individuals to send almost similar requests may occur. As shown in Figure 4, Google's infrastructure was deluged with requests when a magnitude 4.5 earthquake hit the San Francisco Bay area in the middle of the night in October 2019 [50].

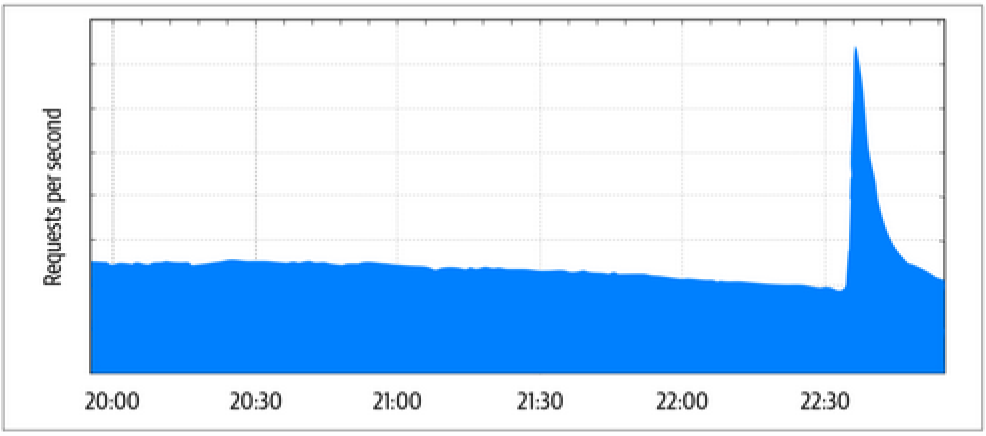


Figure 6 Web traffic, measured in HTTP requests per second [50]

Reliability and safety share many characteristics. All information systems include both as inherent qualities, and it may be tempting to trade one for the other in the name of speed, but doing so will be expensive to rectify. Some of these requirements and constraints may collide, and trade-offs must be made to find the right balance between them [50].

While various design choices offer significant benefits, they may also present certain difficulties. For instance, implementing complex data storage and retrieval mechanisms, ensuring high scalability and performance, and maintaining data integrity can be challenging tasks. Additionally, designing secure and reliable back-end systems requires expertise in encryption, access controls, and network resilience, which may necessitate additional resources and specialised knowledge. Furthermore, the integration of back-end components with existing infrastructure and legacy systems can introduce compatibility issues and complexities [49, 50].

Front-end system design

The front-end design is the second category of design best practices that are looked at in this dissertation. A top-notch front-end design improves user experience and makes it possible for the data to be entered more quickly and accurately [52]. People are less inclined to interact with long-form data collecting that takes time and effort if the design is not compelling or immersive. For instance, employing colour alone can enhance readers' interest in reading by up to 80% [53, 54].

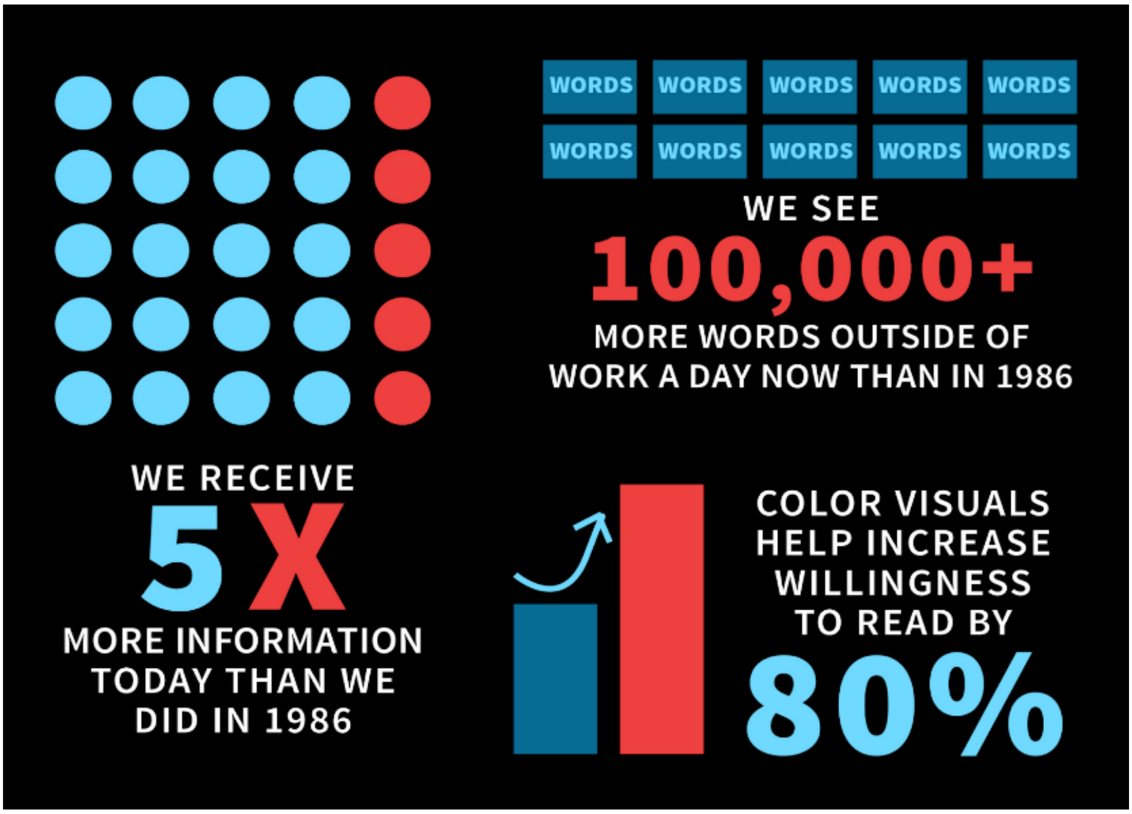


Figure 7: Infographic about user interaction with design [53]

Design becomes critical for a mobile device because of screen size restrictions. For example, in Figure 7, the two studies' information could be conveyed almost seamlessly, although according to Bohn *et al*. [55], this is because of consistency in the infographic's design. The same colour scheme is used throughout.

The rapid spread of Covid-19 in Indonesia prompted the Ministry of Health of Indonesia to develop the Electronic Health Alert Card (e-HAC) application as a measure to mitigate risks. However, users have encountered usability issues, particularly related to menu navigation and the aesthetics of the user interface. The System Usability Scale (SUS) method was implemented to measure system usability based on user feedback, scoring the e-HAC 53.87% in the SUS test [56]. Having such a low score in a user feedback test, Rinaldi *et al*. [56] aimed to address this issue by employing the User-Centred Design (UCD) method to upgrade the interface design of the e-HAC application. The primary objective is to create a suitable User Interface (UI) that enhances user experience. Usability testing, including pre-survey questionnaires and interviews, was conducted to identify and evaluate the problems associated with the e-HAC application. The results of the UCD method implementation and subsequent surveys indicate a significant improvement in the SUS test scores, with acceptability shifting from initially low to an acceptable 85.12%. This demonstrates notable enhancements in the overall usability of the e-HAC application by adjusting the front-end design [56].

Staying consistent can reduce human errors. Prasad *et al*. [48] sort human mistakes into three categories: blunders, slips, and lapses. When tasks are carried out as intended but the incorrect strategy is unintentionally used, blunders arise. A slip is a commission error that happens when simple tasks are carried out without conscious thought. Lapse is a failure to carry out part of a task or the entire task. Slips and lapses are activities that do not involve conscious thought, which might result in omission errors [48]. Meinert *et al*. [57] provide the following recommendations to help users and improve user interfaces:

- Limit the number of data fields displayed on the form
- Limit the number of choices displayed on the form
- Use intuitive data-gathering tools
- Limit exploration until activities are finished

Another method suggested to minimise human errors is through training. Users who are more comfortable with the system are less likely to make mistakes [57]. In Friedman *et al*. [58] two training modalities are suggested: general system training and task-specific training. Users will be able to use systems more successfully and help close the gap caused by unfamiliar jobs thanks to this. Although training has many benefits, a well-designed system will influence how users deal with a particular circumstance. For example, Tomlinson *et al*. [26] noted that immediate data access significantly improves quality. Furthermore, data collection can be guided by implementing advanced skip patterns based on entered values [59].

While adopting intuitive navigation, responsive layouts, and engaging visuals can greatly enhance the user experience, they may require additional resources, time, and expertise to implement effectively. Design choices that involve complex animations, advanced interactions, or innovative features may pose technical difficulties and compatibility issues across different devices and platforms. Moreover, ensuring consistent user experience across various screen sizes and resolutions can be challenging, particularly in the ever-evolving landscape of mobile devices. Accessibility considerations, such as catering to users with disabilities or providing multilingual support, also add complexity to front-end design. Additionally, striking a balance between aesthetics and performance optimisation is a continuous challenge, as incorporating heavy graphics or multimedia elements may impact page load times and overall responsiveness [26, 59].

In this section, the two main focuses of the design used in industries were looked into. The first aspect is back-end design. As seen in this section the focus when designing the back end of a system should be reliability and security. The second focus was on front-end design: while designing the front end of a system, the main factors to consider are ease of use and consistency. In the next section a few examples of how data consolidation has been handled in countries outside of South Africa will be discussed, and how implementing data consolidation has changed.

1.2 Industrial data consolidation in other countries

1.2.1 Aluminium refining industry

René *et al*. [60] contribute to the creation of Life Cycle Inventory (LCI) datasets in this study. This dataset has been tracking the environmental performance of the aluminium industry in a portion of Europe for more than 20 years. Aluminium production involves various processes that consume vast amounts of energy and generate substantial greenhouse gas emissions, contributing to climate change. Additionally, the extraction of bauxite, the primary ore used to produce aluminium, can result in deforestation, habitat destruction, and biodiversity loss. By monitoring and assessing the environmental performance of the aluminium industry, we can identify areas for improvement, promote sustainable practices, and mitigate the industry's adverse effects on the environment [60].

Data consolidation plays a crucial role in this context as it enables the comprehensive collection, analysis, and interpretation of environmental performance data across the industry. By consolidating data from different aluminium producers, policymakers, researchers, and stakeholders can gain valuable insights into the industry's overall environmental impact. This consolidated data allows for a better understanding of the industry's resource consumption, emissions, and waste generation, helping to inform decision-making processes and develop effective environmental policies. By combining data, this study seeks to shed light on the effects that aluminium industrial processes and associated products have on the environment. Their contribution intends to [60]:

- Update the environmental statistics associated with the manufacturing of aluminium cast ingots from waste and scrap aluminium in Europe.
- Provide precise and trustworthy generic industry LCI datasets for the critical steps in the process that are necessary to determine the environmental impact of cast ingots made from scrap and waste aluminium in Europe.

An overview of the report generated through data consolidation can be seen in Figure 8.

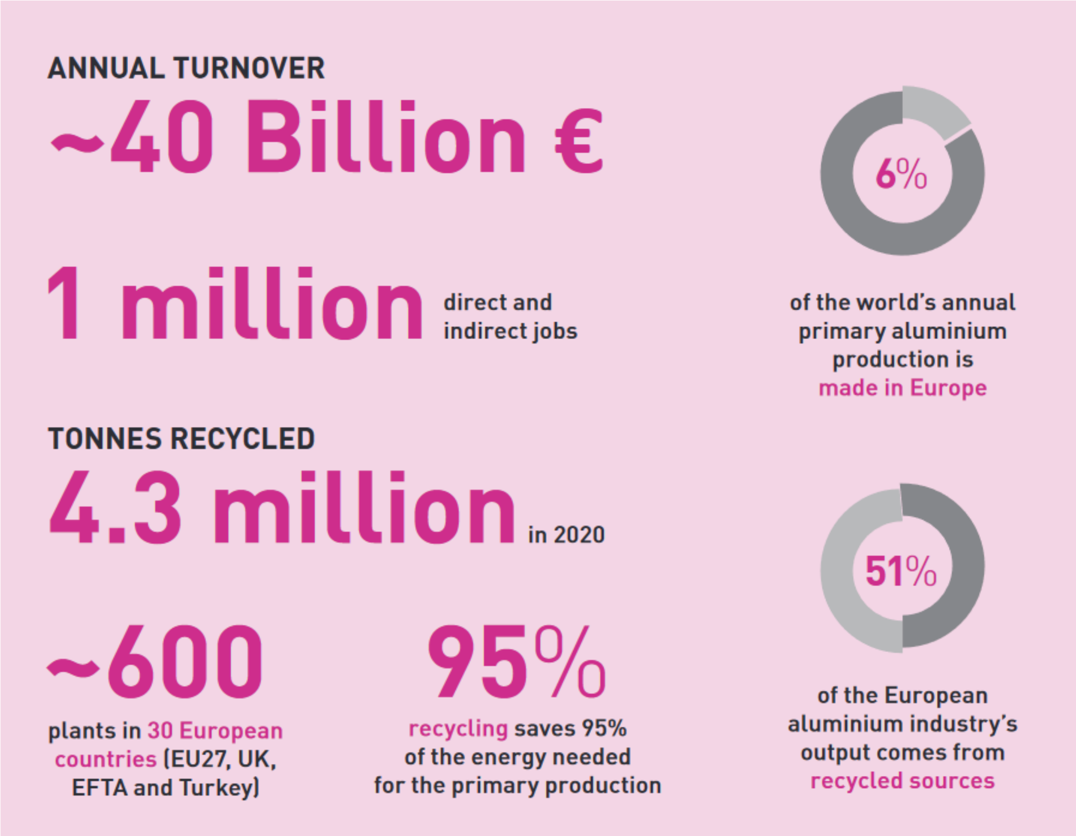


Figure 8: Overview of consolidated data in the aluminium industry [60]

The aluminium industry contributed to the data collection, which is considered representative by industry experts. The collected data covers a significant portion of European production for processes like melting and dross recycling (60%) and salt slag recycling (80%). The collected data was integrated into LCA software (Gabi) to assess the environmental impacts of aluminium processes. Throughout the project, internal experts from the aluminium industry and external consultants from Sphera reviewed and validated the data, methods, and analysis. Lessons learned from previous external reviews were also considered. Although this report focuses on the refining industry and was not externally reviewed before publication, European Aluminium aims to continue seeking external critical review from independent experts for future updates of its Environmental Profile Reports [60]. In Figure 9 the scope of the data collected for consolidation can be seen.

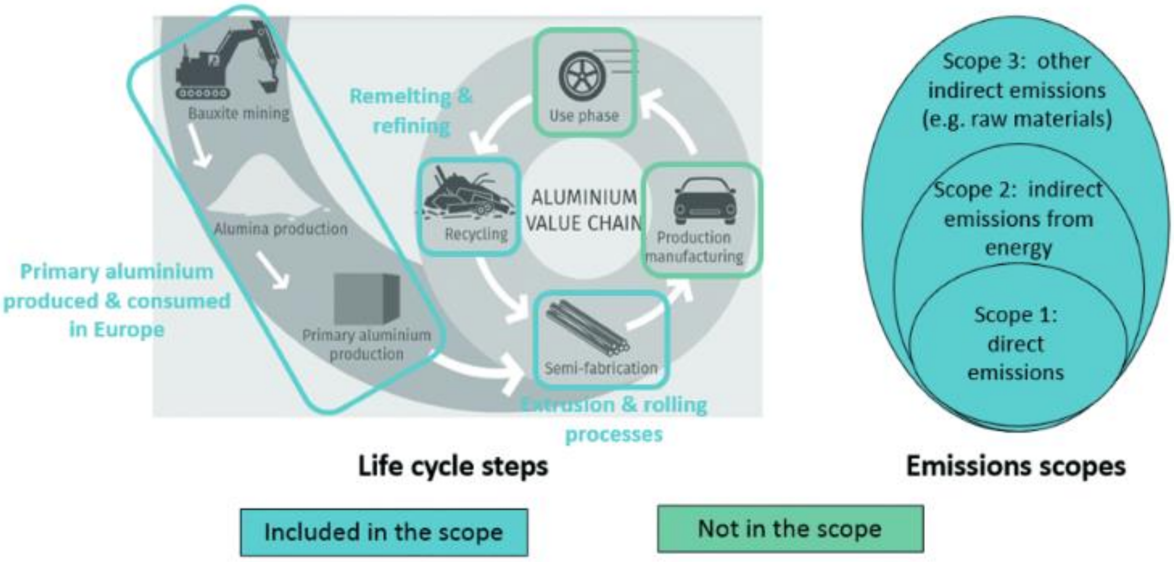


Figure 9: Scope of data consolidation in the LCI [60]

This report provides 7 environmental datasets covering the entire aluminium value chain production in Europe [60]:

1. Primary aluminium produced in Europe
2. Primary aluminium consumed in Europe
3. Rolling mill process in Europe
4. Foil process in Europe (These datasets are prepared in collaboration with the European Aluminium Foil Association)
5. Extrusion process in Europe
6. Remelting process in Europe
7. Refining process

The assessment does not include the production and manufacturing of final products or the use phase of the products, as these datasets are more relevant to individual aluminium customers. However, they will be incorporated when conducting a comprehensive life-cycle assessment of final products that contain aluminium.

Consolidated data have been entered into a software application, in this case, Wufoo, to produce a comprehensive analysis of the environmental impact of aluminium processing. In conclusion, numerous businesses can benefit from the raw materials and energy savings produced by the aluminium refining sector. It serves as a fitting example of a circular economy business strategy that is effective not only for the aluminium value chain but also for other industries [60].

From a societal perspective, the availability of updated environmental statistics and LCI datasets for the aluminium industry contributes to greater transparency and accountability. It allows stakeholders, including consumers, investors, and policymakers, to assess the industry's environmental footprint and make informed choices. With this information, society can hold the aluminium industry accountable for its environmental impact and demand more sustainable practices. It also facilitates the development of policies and regulations that promote environmentally responsible behaviour within the industry, such as setting emission standards or incentivising the use of recycled aluminium. Moreover, the updated datasets have broader implications for society's transition to a more sustainable future [60].

Since tracking of the statistic started, the carbon intensity of primary aluminium production in Europe has decreased by 21% versus 2010 and decreased by 55% versus 1990. The carbon intensity of process scrap recycling (remelting) in Europe has decreased by 9% since 2010. Keeping track of the statistic allows the aluminium industry to improve production to help the environment by not using unnecessary resources [60].

In this section an overview of how the aluminium industry used data consolidation and some of the effects this has had on the industry was discussed. In the next section data consolidation methods applied to radiation research will be discussed.

1.2.2 Radiation research

Being able to consolidate data can lead to a lot of beneficial projects, that would not been able to be done if researcher or individuals were working separately. In the example in section reseachers in the radiation field were able to achieve better results by using a mobile data consolidation application to gather data.

Adverse outcome pathways (AOP) offer a systematic approach to organising and interpreting biological and toxicological data, enabling researchers to identify key molecular and cellular events that contribute to adverse effects. By understanding the underlying mechanisms of radiation-induced harm, AOPs facilitate the development of predictive models, risk assessment strategies, and targeted intervention approaches [61].

Although there is growing interest in AOP in the field of radiation, more research is required to show how AOP might organise biological and toxicological data to help with data interpretation and technique development. In this study, Chauhan *et al*. [61] propose a novel mobile application approach for data consolidation. By leveraging the capabilities of mobile technology, the aim was to streamline the integration of various data sources, such as radiation quality, dose rate, and dose, to enhance the predictive utility of AOP. This approach serves as a crucial step towards progressing research in multiple fields, as it allows for a more comprehensive analysis of data and facilitates informed decision-making [61].

Consolidating data in formal frameworks offers numerous benefits, as outlined below [61]:

1. Tissue level sensitivities understanding: By organising data within formal frameworks, researchers gain a deeper understanding of how different tissues or organs respond to radiation exposure. This understanding helps identify specific sensitivities and vulnerabilities, enabling the development of targeted intervention strategies. For example, consolidating data on radiation-induced lung injuries can provide insights into the mechanisms underlying lung tissue sensitivities and guide the development of personalised treatment options.
2. Finding bioindicators and biomarkers: Data consolidation facilitates the identification of bioindicators and biomarkers, which are crucial for understanding the progression of diseases and enhancing their detection. Through comprehensive analysis and integration of diverse datasets, researchers can identify specific molecular or cellular markers that indicate the presence or development of radiation-related diseases. For instance, by consolidating data on gene expression patterns, researchers may discover biomarkers that can accurately predict the risk or progression of radiation-induced cancers.
3. Making connections between low-dose and low-dose-rate impacts and health outcomes: Formal frameworks for data consolidation enable researchers to establish connections between low-dose and low-dose-rate exposures and their impacts on health outcomes. By integrating data from epidemiological studies, experimental models, and clinical observations, it becomes easier to characterise risks based on disease-based information. For example, by consolidating data on the effects of chronic low-dose radiation exposure on cardiovascular health, researchers can establish associations between exposure levels and the development of heart diseases, leading to improved risk assessment and mitigation strategies.
4. Improving risk and hazard evaluation for co-exposure scenarios: Data consolidation in formal frameworks enhances the evaluation of risks and hazards associated with co-exposure scenarios. By integrating data on multiple environmental stressors, such as radiation combined with chemical pollutants or occupational hazards, researchers can assess the combined effects and potential synergistic interactions. This information is critical for developing comprehensive risk management strategies and establishing appropriate exposure limits. For instance, by consolidating data on the combined effects of ionising radiation and specific chemicals, researchers can determine the magnitude of risks and prioritise regulatory measures to protect exposed populations effectively.

The research areas that were found through consolidating data that can be approached by using the method in the study are outlined in Figure 10:

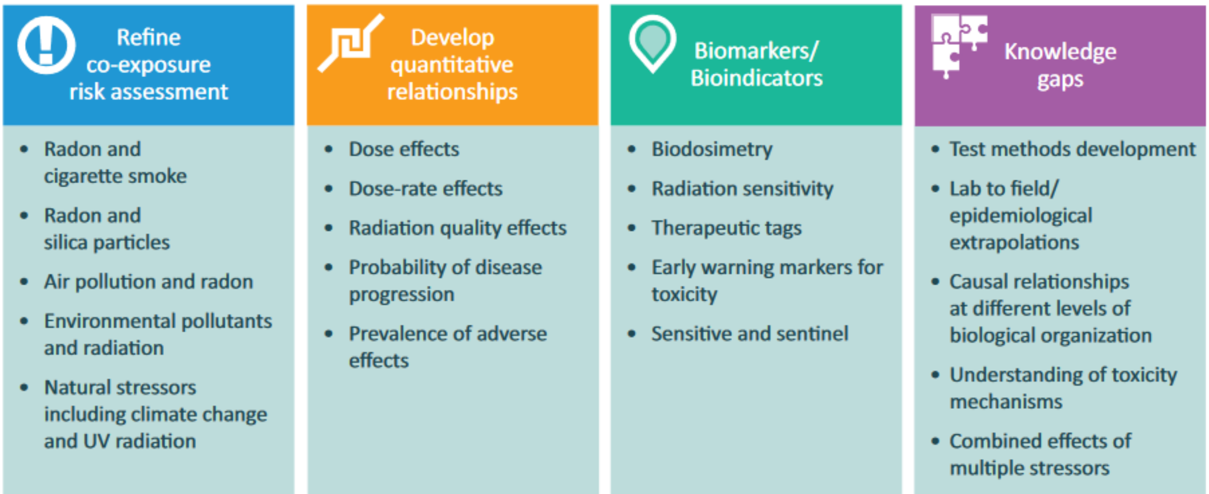


Figure 10: Research areas adverse outcome pathway approach possibly supports [61]

While the application of this strategy has made noteworthy progress in the field of chemistry, this has not yet been the case in the field of radiation. Drawing on and utilising the work being done by the chemical community would be an excellent place to start and help to find similarities and differences to shape the program's inclusion and broader expansion. To make radiation biologists, epidemiologists, politicians, and regulators aware of the AOP program, institutional research work plans will need to include AOP into them through training, knowledge transfer, and outreach programs [61].

The proposed mobile application approach offers a novel and efficient solution for enhancing data consolidation in the field of radiation research. By leveraging the capabilities of mobile technology, this approach enables seamless integration and organisation of diverse data sources, including radiation quality, dose rate, and dose. The mobile application will serve as a central hub for researchers, allowing them to input and access data from various experiments, epidemiological studies, and clinical observations. This streamlined data consolidation process enhances researchers' ability to identify patterns, trends, and correlations within the data, thereby supporting a deeper understanding of radiation-induced hazards and risks [61].

The utilisation of the mobile application approach for data consolidation in the field of radiation research holds several expected outcomes that contribute to advancing scientific knowledge and improving radiation-related practices. Firstly, the mobile application approach is anticipated to facilitate enhanced collaboration among researchers and professionals working in diverse disciplines. By providing a centralised platform for data sharing and collaboration, the mobile application fosters interdisciplinary cooperation, promoting the exchange of ideas, methodologies, and findings. This collaborative environment is expected to drive innovation and accelerate the progress of research in radiation-related fields [61].

The program is well-developed, and the research and regulatory communities for chemical and ecological issues are represented. This knowledge will be used to enhance data consolidation to support radiation hazard and risk assessment. Selected AOP coaches are envisioned as being able to direct projects to produce powerful collaborative proposals that would better transfer scientific findings into societal advantages. Organised international initiatives can promote productive communication between regulatory communities and research experts [61].

In this section a project that consolidated the data in radiation assessment and the effect it could have on the industry going forward were discussed. To implement data consolidation a system that can collect the data and consolidate it effectively will be required. In the next section existing systems that can be implemented will be looked at.

1.3 Review of existing data consolidation systems

A systematic literature review of existing data collection and consolidation systems was used to investigate whether any existing system could be used for data consolidation in the South African industrial environment. The featured that were considered were as follows:

1. Reconfigurable surveys

Data collected with the current configuration should be usable even if the configuration changes after the data is collected. This will save time by ensuring that user input is not required when implementing necessary changes. The solution should be modular in its additions and exclusions when creating and editing the surveys.

2. Basic input validation

This requirement will ensure that fundamental human errors are limited by providing a field filled in when needed and will also ensure that the value in the field matches the expected data type.

3. Validate using other fields

To improve the data quality and the speed at which the data can be collected, the application should be able to validate the current data using collected data. For example, the collected data can filter out options giving the user fewer options or ensuring that it falls within a range of other values.

4. Work offline

With the industrial environment of South Africa, an internet connection is not always a guarantee. This can be because of location and or other circumstances. Therefore, the application will have to work offline for most of its operation and synchronise the data when the internet connection can be secured.

5. Encrypted online server

The data collected with the device can be sensitive. Therefore, the data should be consolidated in a secure online database, and communication and storage in this database should be encrypted to ensure data is not tampered with or accessed without permission.

6. Authorise user with secure login offline

After the user's data is on the device by logging in while the internet connection is available, the user should have a secure way of accessing the data collection and collected data on the device. Out of all the survey and data consolidation systems researched, the biggest concern was collecting the data offline and integrating a secure online database for sensitive data.

The systems found for further inspection are as follows:

1.3.1 Google Forms

Google Forms is arguably one of the simplest form builders available. Users can choose from 12 fields and 9 question types, including text, photo, and video. These forms can then be sent to people via a link to be filled in. Users can retrieve the data on Google Forms and display the data in spreadsheets and charts to analyse the data further. The data inserted into Google Forms is analysed in real time, and the results can be viewed instantly with their built-in support system [62].

Some features are lacking in Google Forms to make it work in the industrial environment discussed earlier in this chapter. The surveys cannot be downloaded offline when an internet connection is connected again. Google Forms also do not have a method to filter the questions or inputs based on data already inserted. The data collected is also stored in their databases, meaning it will have to be retrieved before further processing. Google Forms also has restrictions when used for commercial applications [63].

1.3.2 SurveyMonkey

Another tool for data consolidation found was SurveyMonkey. This tool allows the user to configure forms with 15 different question types. SurveyMonkey also has conditional branching options allowing the user to filter upcoming fields based on responses given. In addition, results can be easily exported into many formats for ease of use. This tool also has a mobile survey template allowing the users to be able to use smart devices as well as computers across various platforms [64].

A study on data consolidation found that SurveyMonkey had limitations on the number of created surveys and respondents. Data export was only available on the paid version of the system. SurveyMonkey has limited offline capabilities, not allowing the user to be authenticated or log in without an internet connection. The tool also does not have input validation per question enabling the form to be configured with a warning to ensure the user is inserting the correct data. The setup of surveys is complex and might require additional training to set up on the user side [63].

1.3.3 Microsoft Forms

Another data consolidation system considered was Microsoft Forms. This system uses a drag-and-drop editor to configure forms. Microsoft Forms can export data collected to the system using Microsoft Excel. This tool allows a calculation to be embedded into the state, allowing users collecting data to run a preconfigured analysis and help with data collection. Like WordPress, many different plugins can be inserted to ensure data is collected in the way that the user wants. The form page add-on can also be installed on Microsoft Forms, allowing multi-device support and configuration for each survey [65].

Microsoft Forms' extra features can get complex to use, which may require training to ensure the user can utilise all the functionality. In addition, Microsoft Forms cannot log users in while offline, creating problems in the South African industrial environment. Data collected will have to be extracted using Excel or other formats and then moved to an existing database if further processing is required since Microsoft Forms does not have a secure communication method with third-party databases [66].

.

1.3.4 SurveyLegend

SurveyLegend is a next-generation online form builder that is exceptionally mobile-friendly and allows users to create and take surveys in many ways. This tool enables sharing of the user's surveys via email, social media, and SMS. The survey colour scheme can be edited to ensure it aligns with the current company setting up the survey, including a logo that can be added. In addition, SurveyLegends can be used to analyse the data using a built-in system that will work automatically as soon as data is synched [67].

This tool does not allow input fields to be configured using user input into a previous slot. This can slow down the data collection process and make it hard for the user to follow the survey. The data-collection fields also do not allow for different data specifications while the data is being collected. SurveyLegend also does not allow unprocessed data to be stored and analysed in third-party servers, and it restricts the number of surveys created [63].

1.3.5 DoForms

DoForms allows companies to easily capture information on their smartphones and other mobile devices by creating a mobile form workflow solution. The application also allows the users to build forms with a drag-and-drop interface and then send the built form to other devices. In addition, DoForms provides data validation, calculations, and user management [68].

A study done with health workers in rural areas found that using DoForms can potentially increase efficiency in data collection since data can be accessed in real-time to resolve immediate edits. However, DoForms required extra time and training for the health workers to use it efficiently. DoForms did not meet the requirements due to the limited database and offline use functionality. DoForms does not support integration with existing databases that are not on their system, as the application's available online database is not isolated. In non-isolated systems, it is easier for other members to use the system to gain access to information. Due to this, users are unable to store sensitive data. The application allows limited offline data storage, meaning that very few calculations can be done on the device during offline inspections [69, 70].

1.3.6 FlowFinity

FlowFinity uses a point-and-click interface to allow users to create interactive dashboards to collect and store data. This will enable users to create mobile applications without coding knowledge. This functionality means businesses can save time and money by replacing paper and spreadsheet systems with applications [71].

The interface used in FlowFinity is unsuitable for this project as the application allows users to create custom interfaces that may require additional training to be administered per inspection form. In addition, FlowFinity does not differentiate between data collection types which will be needed for validation [70, 72].

1.3.7 Wufoo

Wufoo serves as a cloud-based form builder that aims to simplify creating forms. Wufoo has a collection of over 400 templates to choose from. This makes the creation process very customisable and allows users to create forms with the correct look and feel. Wufoo also accepts HTML code to allow for further customisation. This data consolidation platform allows conditional logic to collect less redundant data, has encrypted communication with an online server and can inform users of the results via email [73].

The limitation of the number of created surveys and survey questions was also identified. Wufoo cannot change surveys after data has started to be collected and still uses data with a new set. In addition, mobile implementation is limited, and collecting data offline is restricted. Finally, the form configuration must be manually saved, creating extra steps to get the data from the forms [63].

1.3.8 Formitize

Formitize creates a paperless form system by using applications and connecting the cloud to store data. The application allows for form building, distribution, and accessible data collection. Furthermore, Formitize offers offline support and secure data management. Additionally, the application allows for data validation and communication between the devices currently working on the application [74].

Maddox *et al*. [75] collected data for a study using Formitize and found that existing data collection methods provided limited and unreliable information. As a result, the data from the first three months was recoded to fit the new fields. The entire process of creating and implementing the application needed a lot of time, resources, and strategic planning, including choosing a vendor, testing the functionality, defining what data would be collected, choosing how to analyse it, and training personnel. To guarantee that the application produced useful data to minimise future revisions, this expenditure was essential. Since the launch of the application, the amount of programme data collected has increased significantly and steadily. For example, the number of reported people linked to mental health services was 391 in the six months before the application was launched, while it was 929 in the six months after the application was launched [75].

Formitize has limited offline capabilities that do not allow the application to validate data based on historical data. This application only has secure databases in other countries. Without a secure database, users cannot store sensitive data. Formitize also cannot persist in changes to configurations for previously collected logs. The editing of records collected of configurations that have been changed after the record was collected is impossible on Formitize. This means the user cannot access logs collected after a configuration was changed [70, 75].

1.3.9 ClickUp

ClickUp is a project management software that is mainly very customisable, featuring many functionalities from task tracking, chat windows, automation of tasks and other management features like planning software. The data consolidation side of ClickUp is more aimed toward project management allowing users to set up data collected to be used to forward tasks or create bug reports. ClickUp will enable users to import data from various sources, including other management software, but has limited functionality when exporting, not allowing API or other automated exporting [76].

For the postgraduate program at the University of Pretoria, a redesigned curriculum has been in place since 2014 that allows students a hybrid learning environment. Students received one face to face class per week, and other classes were done independently and needed active student involvement in ClickUp. This innovative method of online participation encouraged students to study and actively engage with the course materials, which helped them be more prepared for their assignments, which were also turned in via the course management system (ClickUp). Some students recommended further instruction in the usage of the online course management system because they felt that these activities were "overwhelming" and "too much" (ClickUp) [77].

This system has applications for devices including mobiles, desktops, and smartwatches. However, ClickUp does not have the necessary offline capabilities to collect data in industrial environments. This system will also require training to teach users the various aspects and functionalities that can be used with this system [76].

1.3.10 ProntoForms

ProntoForms allows for the creation of customisable forms that can be used on mobile devices to collect data. An easy-to-use, web-based tool with a drag-and-drop, no-code visual application-building method is offered by ProntoForms for non-programmers. Its platform is best suited for automating paper forms, developing new applications in prototype form, and updating or mobile-optimising legacy desktop-bound software. This application has support for both online and offline data collection. ProntoForms makes it possible to analyse, track and share data collected with devices with the application. ProntoForms' custom creation allows users to create their forms through their system [78].

Since ProntoForms' custom creation tool can only create one survey at a time, the creation process will have to be repeated for every possible survey that needs to be made. This can be tedious as many surveys are similar with only one or two differences. In addition, ProntoForms' log collection also does not allow input validation, meaning invalid data might be collected [66, 78].

1.3.11 MICAVS

MICAVS allows collecting data online and offline, connecting to a remote server and storing and synchronising all data through all applicable devices. The application logged in users using a key generated for each device, allowing data to be traceable. A significant strength of this system is its ability to store logs and their specific rendition of the back-end data, allowing for the manipulation of the data even after the back-end data and configuration have changed [70, 79].

This application was not created to collect data for a two-dimensional interface like a table but for one-line survey purposes (the client requested that this feature be added). As a result, the application could take surveys in table format but at the cost of the user's interface. The current user interface takes substantial time to set up a survey entry to take in information [70, 79].

MICAVS`s synchronisation task takes too long to complete. The synchronisation task includes parsing and storing received data. There is room for improvement with this task. In addition, the legacy application was written for Android, excluding users with devices running on other platforms [70, 79].

1.4 State-of-the-art summary

Using the information gathered in this chapter and the pros and cons discussed in the previous section that looks at existing consolidation tools, several requirements needed for a data consolidation that will work in the industrial environment of South Africa were identified. The requirements identified were as follows:

Table 3 provides a summary of which requirements the systems that were investigated met.

Table 3: State-of-the-art matrix

Sources	Reconfigurable surveys	Basic input validation	Validate using other fields	Work offline	Encrypted online server	Authorise user with secure login offline
Google Forms						
SurveyMonkey						
Microsoft Forms						
SurveyLegend						
DoForms						
Flowfinity						
Wufoo						
Formitise						
ClickUp						
ProntoForms						
Micavs						

	Did not satisfy the requirement
	Satisfies requirement

No feasible solution for the problem stated in the first chapter was found in the survey of data consolidation systems. Therefore, a system must be created to meet all client requirements.

1.5 Problem statement and objectives

The harsh industrial environment of South Africa prevents companies from collecting data or digitising existing data collection systems. A lack of resources and training compounds this problem. Having a curated mobile application to collect data while addressing the challenges in the South African context could make it faster and more convenient for companies to collect and analyse data.

Consequently, a need exists to develop a data collection application that can methodically collect traceable, high-quality data within the South African industrial environment.

To determine the possible benefits and shortcomings of implementing a curated data collection application in an industrial South African system, the following objectives need to be met:

1. The data must be collectable in a variety of different configurations and for different scenarios.
2. The application should also be able to work offline to ensure that it is usable in an industrial environment.
3. Using previously collected data and current information on how trustworthy the data being collected is should be determined.
4. The application should also give the user feedback if the data currently being taken is possibly untrustworthy. This could lead to less human error, increasing the data quality. Harsh environments hamper the accuracy and readability of manually collected data. This results in data becoming untraceable and unfit for potential financial rebates.
5. The configuration of data collected should also be able to change while still maintaining the ability to edit data that was collected before the change in configuration.
6. The application should have the ability to identify the user using the device even when not connected to the server to ensure that data is tracable.

1.6 Dissertation overview

In this section, a background to the economic circumstances facing the South African industrial sector was given. Optimising systems in the South African industrial sector could help mitigate some of these challenges and a possible solution to optimising these systems was given. This solution requires good-quality data.

Feedback on the methods that could be used to get good-quality data was provided. The data collection method that was found to be reliable and mitigate errors while collecting data securely and efficiently was the mobile approach. The method of storing data is reliable and cost-effective while mitigating errors caused when getting the data ready to be used as an online storage method. The main design elements to consider when looking at back-end design are reliability and security and when, looking at front-end design, ease of use and consistency.

Existing systems were scrutinised, and their strength and weaknesses were listed. No current system was found that met all the specified requirements. Working offline was a challenge for most systems. The system needed to be able to log a user in securely without having access to an online server, and then sync the data when there is an internet connection. Having a modular survey creator that could add or remove fields and edit data types of existing data fields to allow for ease of use when designing the data collection layout and using the system to collect the data is another requirement that was found.

In the next chapter, the design of the method used to reach the study's objective is laid out. Then, the design and implementation of the front- and back-ends of the system are discussed. Finally, after the artefact development is laid out, the program's implementation method and the case studies for testing the methodology are discussed.

In chapter 3, the design from chapter 2 is implemented in case studies. The case studies and data handling in these cases are discussed in detail. The system will then be verified using the data gathered in the case studies. The results of implementing this solution in two case studies will then be given.

Lastly, in chapter 4, a conclusion will be drawn about the effectiveness of the proposed solution and the possible follow-up research that can be done.

2 DESIGN OF THE DATA CONSOLIDATION METHOD

2.1 Introduction

2.1.1 Design overview

The agile approach was used for the project to solve the engineering problem. In Figure 11 an overview of the agile approach can be seen. This approach starts by defining the requirements for a project by performing research and analysis. The next step is to design the necessary user interfaces, system interfaces, network requirements and databases. The next step is to develop the designed solution and test if the solution meets requirements. As shown in Figure 11 if the solution does not meet requirements the project moves back to design phase. The agile approach goals are defined as increasing the complexity of the solution until all requirements are met. This part is completed in this chapter.

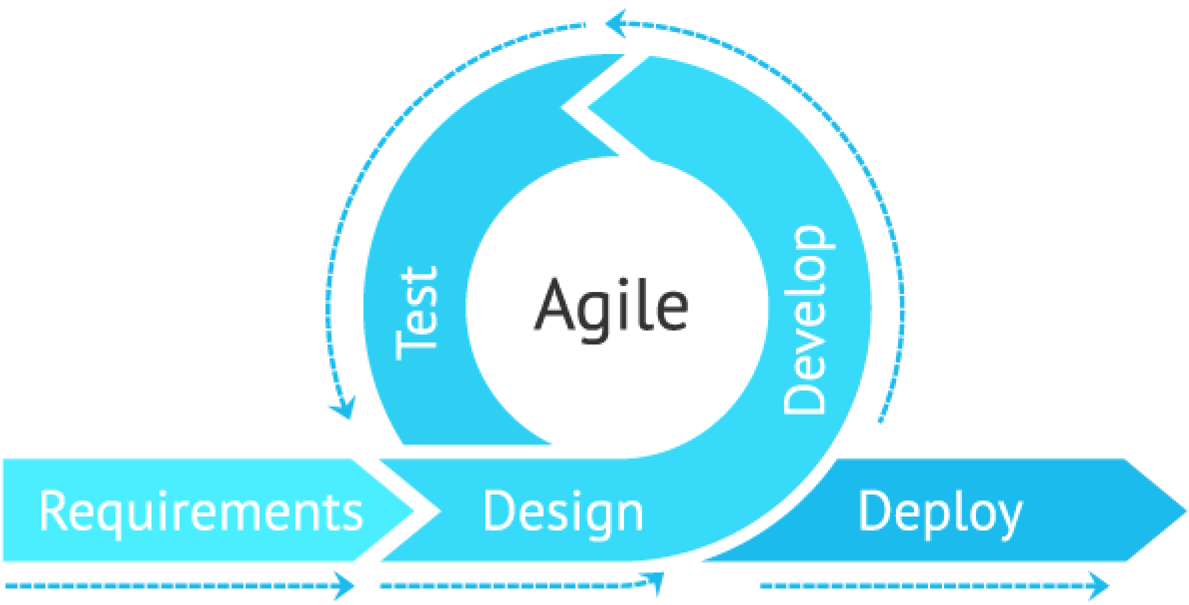


Figure 11: Overview of the agile approach

In Figure 12, an overview of the system's architecture can be seen. The system can be divided into three main segments. Firstly, the application will be the primary interaction for the users collecting information through the system. This part will have to run on rugged devices and store the data locally until it can be synchronised with the rest of the system using an internet connection. This part includes the user interface (UI) and the associated UI-driven functionality. Secondly, the setup and API system will be where the configuration for collecting data will be done. This part will also facilitate communication between the mobile application, the setup, and the back end.

Finally, the support system part of the artefact will store the information gathered by the system and allow users to view and retrieve the collected information.

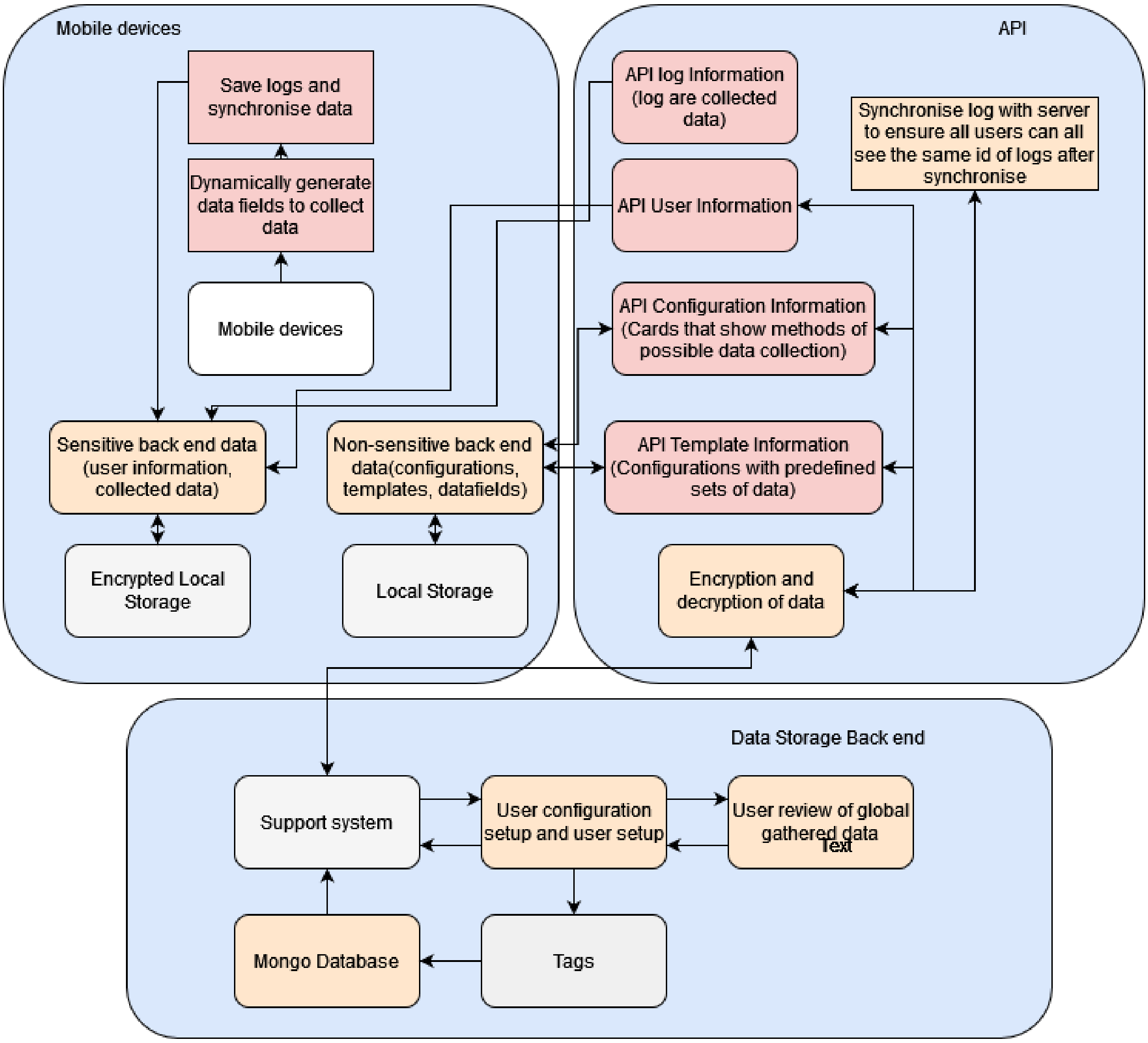


Figure 12: Mobile System Architecture Overview

The devices selected to run the application will have to be rugged (heat, water and drop resistant) enough for the South African industrial sector. The selected device will also have to be able to accommodate the necessary inputs to complete the data collection tasks, which includes a camera, GPS, and internet connection. The given devices should also be able to run the application. This means the hardware and software of the device should be adequate, including the iOS or Android operating system, and have a decent amount of onboard storage to store the data collected.

The application will have to work on the devices used in the South African industrial sector, including older and slower devices with different screen sizes.

2.1.2 System design criteria

An industrial environment that already makes use of an existing data collection and consolidation system will be required to assess and ultimately prove that mobile application systems improve data consolidation. Once a system in which data is already gathered and stored has been identified, a new system can then be implemented, and the effects can be measured by the new system. Another requirement was that of personnel; personnel that already performed data collection using existing systems, albeit paper-based, was essential to the success of the project. The primary verification of the system being improved is if the case studies report monetary value from using the system.

Systems in the industrial environment were chosen to apply the methodology. These systems had personnel already collecting data but had shortcomings such as the data not being able to be used or in a problematic format that requires extra time to be processed, for example. These users can then be asked to use the system without adding extra personnel. By using the infrastructure and personnel already in place and replacing the data consolidation system with a mobile solution, factors such as additional time being spent on collecting data, or additional resources being spent on personal that collect the data can be eliminated. If the users then report monetary value from using the new consolidation system, the value of the new system can also be validated.

As seen in Chapter 1, paper-based logs create logistical problems from storing data to retrieving it and reusing or repurposing it at a later stage incurring possible errors because of this method. A SCADA and other permanent data collectors have a significant upfront cost and cannot be used to collect spontaneous data. Finding existing systems that are data-heavy can be improved by using mobile hardware, such as cameras and recordings in addition to the existing tools.

Using the properties discovered during the literature study for a data consolidation system, several criteria will need to be addressed by the chosen method. The typical work sequence of mobile data collection systems with consolidated data was described in 1.1.5 [36]:

1. Design forms for data collection
2. Give out forms to field devices
3. Collect data with mobile devices
4. Submit data to a centralised database
5. Process data as it is received
6. Provide results and reports

Using this work sequence as the basic flow of the chapter ensures that the method can do the basics of data consolidation applications. Using existing digital consolidation systems described in sections 1.3 and 1.4, the main properties found to be significant were as follows:

1. Reconfigurable surveys
2. Basic input validation
3. Validate using other fields
4. Work offline
5. Encrypted online server
6. Authorise user with secure login offline

Using the steps for data consolidation along with the additional criteria discovered through literature, the method shown in this chapter was designed.

The system developed to collect data sets up the dynamic configuration of data fields. This allows a unique data collection method to be created for each user of the application. Since the first objective of a case study would be to make the configuration used to collect the data that will be consolidated, the steps needed to set up these configurations for the user will be discussed in this section.

2.2 Support system development

2.2.1 Configuration overview

The first step is creating data collection forms, and this involves knowing what data the user is going to require. Each data piece the user will require is then translated into data fields. These data fields are then linked to the configuration. The last step required is linking the created composition to a user, giving a user access to that configuration.

The workflow created for this method to create and store data collection forms can be seen in Figure 13 below. This workflow was created keeping the first two criteria in mind, namely reconfigurable surveys, and basic input validation. To ensure the surveys are reconfigurable, each element should be able to be added and removed separately and then linked.

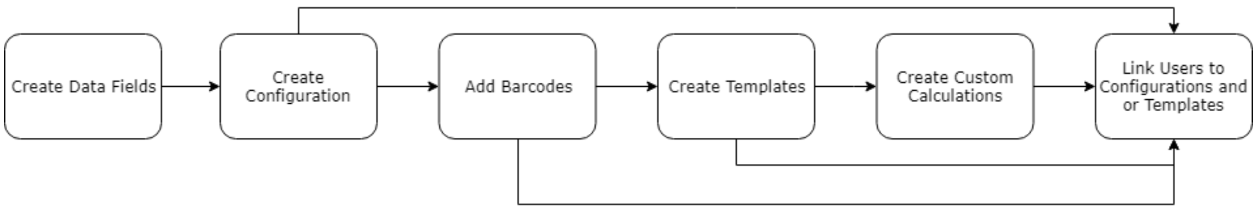


Figure 13: Configuration creation diagram for required data collection

Once data fields and configurations are established, the configuration can be linked to a user. Another optional step is linking extra calculations and templates to the data fields. This allows another form of validation to be built into the program that will be able to warn the user and avoid user errors: this coincides with criteria 3 which is to validate using other fields.

2.2.2 Data field setup

Configuring the data collection and data structure that the user will require will involve taking the user's data needs and creating data fields for each data set required. Then, each data field can be used to collect the data. Figure 14 below shows the system with generic data fields added.

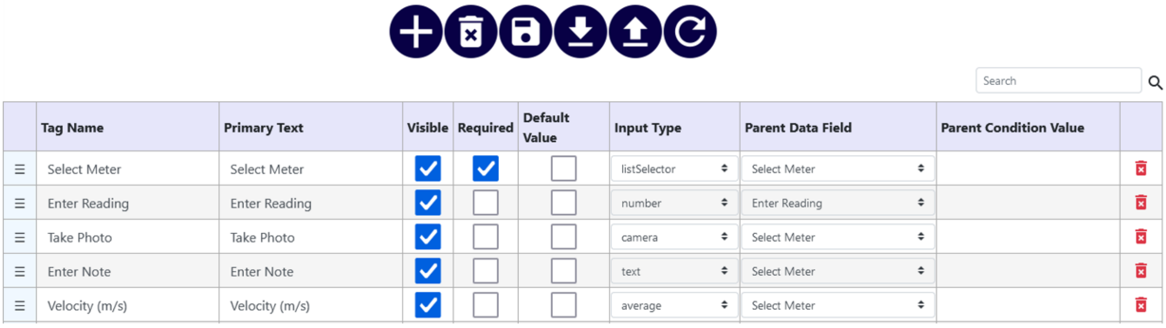


Figure 14: Configuration screen for individual data field setup

Keeping the first three criteria in mind, namely that the surveys should be reconfigurable, the data should have basic input validation and the data should be able to validate using other fields, the following input was chosen for the data field:

1. Tag Name: Unique identifier used to store the data field.
2. Primary Text: The primary text will display on the application and guide the user on what to fill into the data field.
3. Visible: Should the data field be visible to the user.
4. Required: This is the first of the conditional collection settings. By using this, a data field must have a value to continue to add data to the next field. This ensures that the user will not be able to skip essential data.
5. Default Value: The value inserted into the field marked as default is saved. The next time a default field appears to the user, the last filled-in data will appear, allowing the data to be collected more efficiently.
6. Input Type: By choosing the input type, the application will generate a view and database storage location to collect the data type selected. The complete list of input types will follow when discussing data collection. This will also be used for basic input validation, the second criterion.
7. Parent Data Field: This value is used to link multiple inputs together for assorted reasons, such as:

7.1. if another input should filter the list

7.2. if this is a value stored on a photo being taken

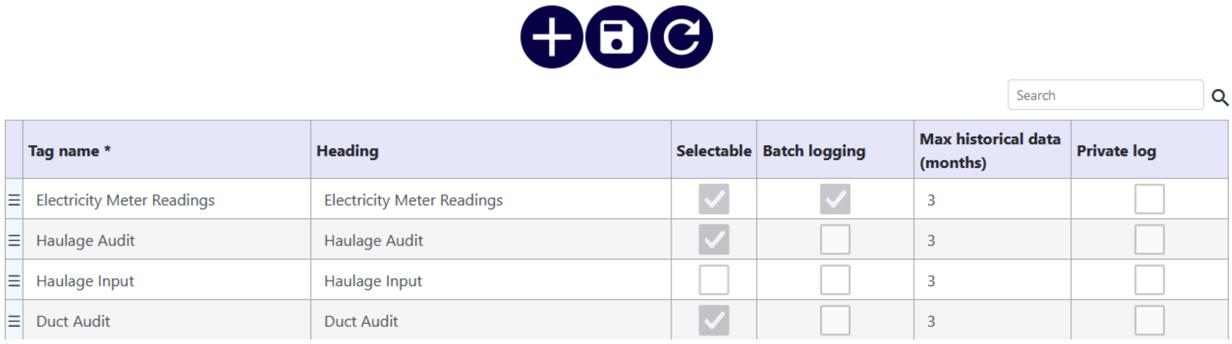
7.3. if this field should be visible if another input is filled in

This is another conditional collection setting. This allows data to be collected faster by filtering what is gathered or adds redundant information by linking fields together, such as connecting a photo input field to a number field. These related data fields can then be used later to verify that the data inserted in one matches the other, making data harder to falsify.

8. Parent Condition Value: This is used to make a field invisible if the condition inserted is not met. It allows additional filtering to enhance data collection and avoid confusion for the user.

2.2.3 Configuration setup

To ensure the survey is reconfigurable, the data fields should be linked to the survey, and a configuration should be made. The configuration is a collection of data fields in a specific order. In Figure 15 below, the table used to create the configuration can be seen with some generic configurations.



Tag name *	Heading	Selectable	Batch logging	Max historical data (months)	Private log
Electricity Meter Readings	Electricity Meter Readings	✓	✓	3	☐
Haulage Audit	Haulage Audit	✓	☐	3	☐
Haulage Input	Haulage Input	☐	☐	3	☐
Duct Audit	Duct Audit	✓	☐	3	☐

Figure 15: Configuration creation screen for grouping setups

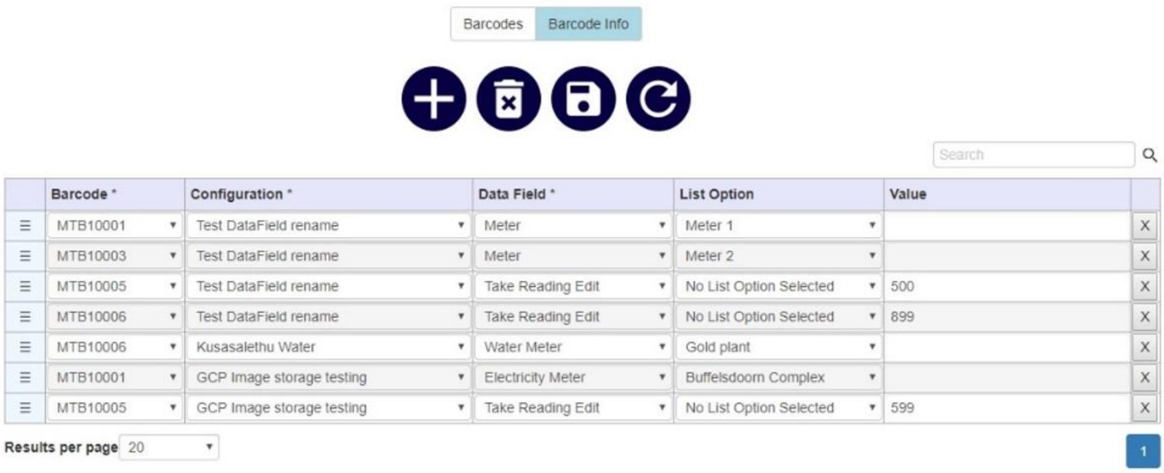
The fields chosen for a configuration to ensure it displays correctly on the application and the server and allow users to use the configuration were as follows:

- Tag name: Unique identifier used to store the configuration.
- Heading: The description of the collection that will be shown on the application.
- Selectable: This property is used by the application to allow users linked to the survey to be able to open the configuration or not.
- Batch Logging: This setting can be enabled if the same configuration is used repeatedly. This allows the user to collect the same data type without choosing the configuration again.
- Max Historical Data: This setting allows the user to specify the months the data collected should be stored on the device and online.
- Private Log: This setting allows the users to specify if only the user who collected the data can see the data or if all users' collections of the same configuration can be seen. This adds an extra level of privacy if the data should not be shared.

By completing the setup of the configuration, the user linked to the configuration can synchronise the application and use the configuration to collect data. The following sections allow extra functionality to be added to the configuration to make the data collection process more efficient and the data collected more reliable.

2.2.4 Barcode management

The application also allows for barcodes to be added to the system. Scanning a barcode on the application allows extra information to be inserted automatically. The data can also be used to navigate the user. The first step would be to add each of the barcode IDs. After each barcode ID is inserted, the info linked to the barcode can be added. Figure 16 below shows the table used to create barcode information with some testing barcodes.



	Barcode *	Configuration *	Data Field *	List Option	Value	
☰	MTB10001	Test DataField rename	Meter	Meter 1		X
☰	MTB10003	Test DataField rename	Meter	Meter 2		X
☰	MTB10005	Test DataField rename	Take Reading Edit	No List Option Selected	500	X
☰	MTB10006	Test DataField rename	Take Reading Edit	No List Option Selected	899	X
☰	MTB10006	Kusasalethu Water	Water Meter	Gold plant		X
☰	MTB10001	GCP Image storage testing	Electricity Meter	Buffelsdoorn Complex		X
☰	MTB10005	GCP Image storage testing	Take Reading Edit	No List Option Selected	599	X

Figure 16: Barcode Info Grid

To ensure that the barcode stores the relevant data required to open a configuration and fill the linked information into the application, the following fields were chosen for the setup of a barcode:

- Barcode: The ID of the barcode and the information should be linked to it. Linking more than one barcode ID to information will cause the application to give users options as to what should be collected.
- Configuration: The configuration that the barcode is linked to.
- Data Field: The data field that the barcode can fill with data.
- List Option: If the data field type is a list option, a list option can be selected to be preselected on the data field.
- Value: If not a list selector, a number or a string can be inserted here for the barcode and scanning the barcode will store this data in the selected data field.

When a barcode link is created, it can have either a list option or a value linked to it. This data is used to prefill that configuration based on where or what the barcode is. This can save users time by giving them fewer fields to fill in. If the barcode being scanned is related to a list option, the data field collection page will open, and the data field associated with the linked list option is auto-filled with the linked list option value. When a value is inserted into the scanned barcode, the data collection page will open, and the related data field will be filled with the pre-configured value. When more than one barcode info is filled in for one barcode ID, the application will show both options to the users and allow the user to choose which configuration is to be filled in.

2.2.5 Linking barcodes and data fields

After creating the data fields and barcodes, the configurations must be linked to the layout. This allows the survey to be edited easily and the data to be collected in the precise order and structure the user requires. For example, in Figure 17, the view that is used to add data fields and barcodes to a configuration can be seen. By clicking the add button under the data fields or barcodes category, the possible data fields or barcodes can be selected. Once selected, dragging and dropping data fields or barcodes will set the order of appearance for the user in the application.

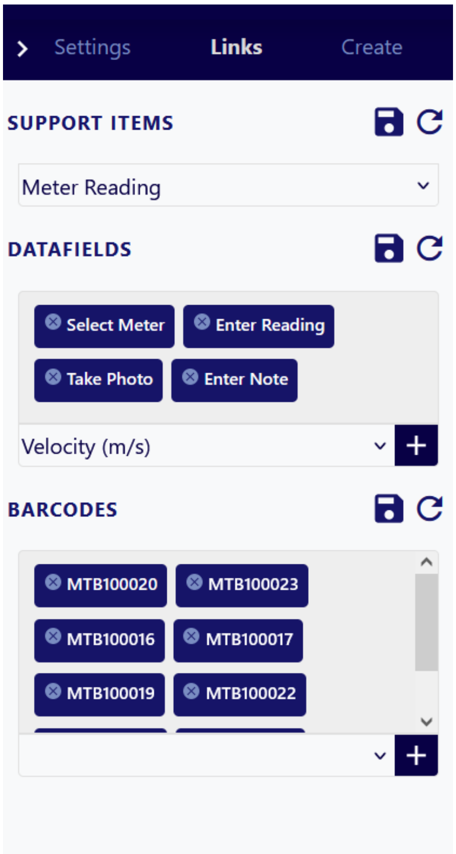


Figure 17: Linking settings for configurations

2.3 Application development

2.3.1 User interfaces

The second step in a data consolidation application is to get the configuration to field devices and step three is collecting data using mobile devices. In this section, a mobile application that can use the structure setup in the previous step and can be used to collect data will be shown. An overview of the application can be seen in Figure 18.

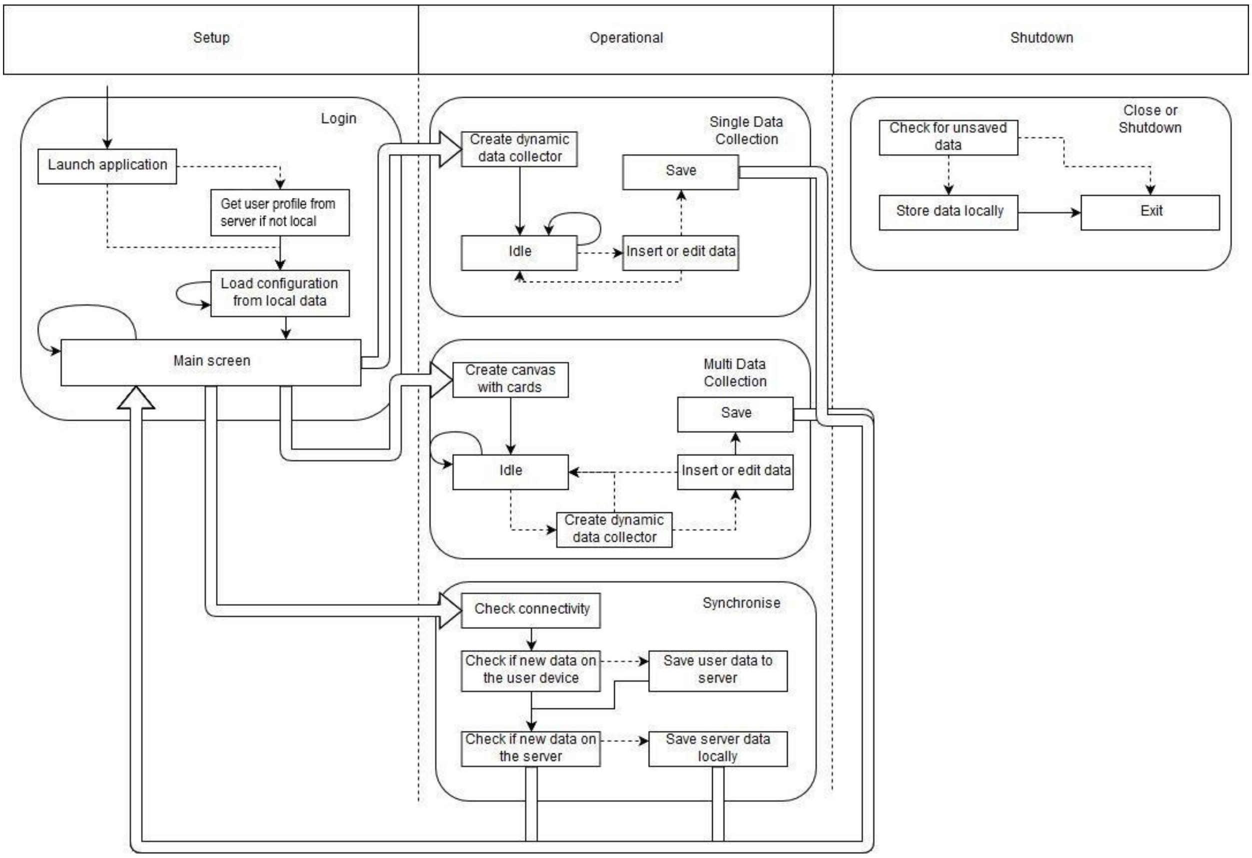


Figure 18: States and Modes diagram application overview

Following the design parameters in section 1.1.6, the user interface is designed to have minimal tasks on the screen at any given time. This allows users to use the application more quickly and make fewer errors [55]. The user can create new entries by collecting new data or editing the old ones in the operational state. The user will also have the option to synchronise the data on the device with the server if the device has a connection to the internet. Dynamic changes are saved locally when the application is closed to ensure data persistence.

2.3.2 Support services

When introducing a new system into an industry, the workers will not necessarily know how to use the system. Therefore, the workers must be trained to use the devices and mobile applications. Training the workers on using the devices and the applications ensures two essential factors. Firstly, the data collected is traceable, plugin reliable and consistent; secondly, the data is managed more efficiently [13, 15].

Having personnel that can collect reliable data efficiently will increase the chances of the success and adoption of the project. Personnel will have to be taught how to collect data with the device and synchronise the data, as discussed next.

The user interface will also be able to collect several types of data while staying consistent in the approach to collecting data. The user interface also needs to change depending on the configuration setup for the specified task.

When creating a new entry or editing an existing one, a dynamically generated page where data can be collected is displayed. This page creates a unique data field for every indicated data type and a storage location for each datatype within a configuration.

In Figure 19, a diagram explaining the process is shown. The first step is to get the configuration ID. This is then used to retrieve the configuration from storage. With the structure in JavaScript Object Notation (JSON), a loop will go through all the data fields where, for each data field, a component on a form will be generated.

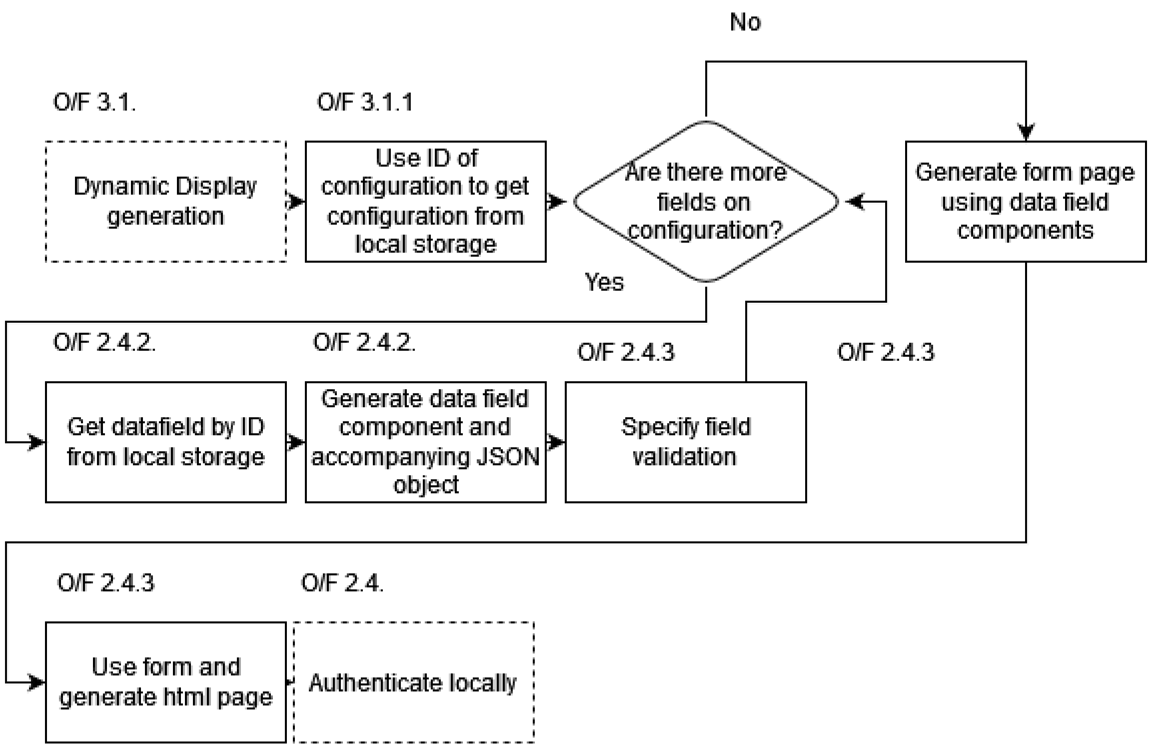


Figure 19: Functional flow - O/F 3.1 Dynamic display generation

For each component on the form, a field in local storage will be created to house the necessary data collected by the data field. Once all the form components are dynamically added and the data structure to store it is generated, the form is then added to an HTML page and is displayed to the user. Being able to generate the display from the stored data helps to achieve criteria four of working offline since the system does not need any new data from the server to generate the forms.

2.3.3 Local storage

In the South African industrial environment, there is often an untrustworthy internet connection or no internet connection at all [79], which means the device will not be able to communicate with an online server while in operation. Therefore, data must be stored on the device to allow the system to work in industrial environments where no internet or other communication mediums can be accessed. Being able to store the data locally before synchronising the data allows the application to work offline which satisfies criteria four as outlined.

Devices with storage systems near capacity or at capacity will slow down or not function properly. This is because the data stored locally will have to be cleared to ensure that the application does not use too much space on the device's hardware and still allows the device to run efficiently.

To ensure the confidentiality and security of the data, the data stored on the device should also not be accessible unless using the mobile application. Using encryption on the data stored on the device will allow only the applications which can decrypt it access to the data.

The entered password is hashed with stored salt if online access is unavailable, and a local user exists. Salt is a randomly generated string added to the password before encrypting. This hashed password is then checked against the stored password. The user is logged in if these match, as seen in Figure 20.

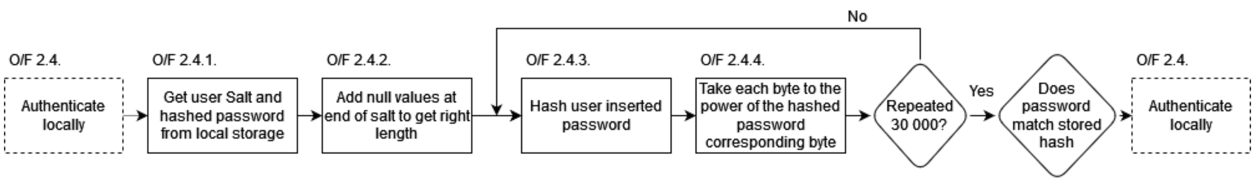


Figure 20: Functional flow - O/F 2.4 Authenticate locally

Next, the password is hashed by Secure Hash Algorithm 512 (SHA512) encoding, which is a hashing algorithm that returns a fixed-size string. In addition, the user's salt string is stored locally for an extra layer of security. Each byte in the array is then taken to the power of the corresponding byte in the previous hash. This process is repeated 30 000 times, and the results are matched with the password stored locally. Being able to login in the user securely while offline satisfies criterion six of the project authorising the user with secure login offline.

2.3.4 Log creation and extended processes

When creating the log, several factors must be saved for liability, e.g., name (of an authenticated user), date, location, photographs, or other corroborating data. This ensures that the data is trustworthy and can be checked later. In addition, the log creation methods ensure that this data is collected and saved alongside the data, and this will help ensure the reliability of the data collected using the system.

When the user chooses the new entry options, the program opens the cycle described in Figure 21. The cached entries for the user are loaded and displayed. There exist two possible entries that the user can complete. If the user chooses an entry described as a single-entry mode, the user starts to collect data. If the choice is a multiple-entry mode, the application opens a secondary window where data can be entered repeatedly. Once the data collection is completed, the data is checked to see if it meets the requirements set by the system. If the conditions are met, the data is saved locally.

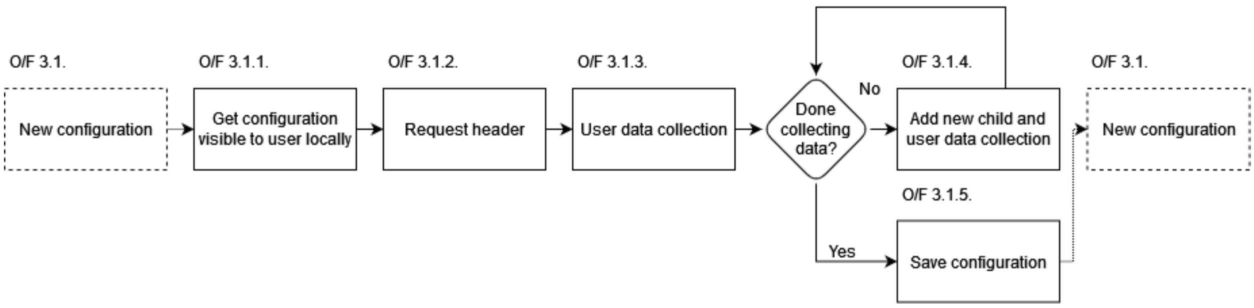


Figure 21: Functional flow - O/F 3.1 new entry

If the user chooses the edit entry option, the cycle seen in Figure 22 is opened. The applicable collected entry is gathered from the local database and displayed to the user. The user then chooses an entry to edit. When the information is saved, the requirements of the data are checked. If the check is successful, the data is saved locally.

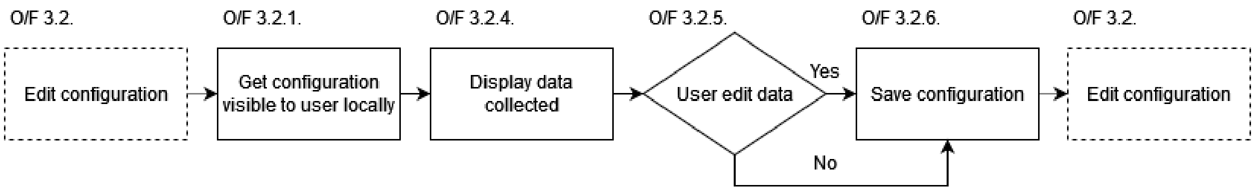


Figure 22: Functional flow - O/F 3.2 edit entry

The data of each case study is in a unique configuration as set up by the users to meet their requirements for the data. This configuration is then dynamically translated by the application: a front-end to collect the data and a backend-end to store the data is created for each specific case study. This also means that the data returned to the server will be in a particular format created by the application. The data structure will also be created and stored in a log to ensure that the data can be translated when synchronised.

In Figure 23 below, the application's main screen can be seen after a user logs in. The application’s users could then select either to scan a barcode to start the data collection process or to capture data which would allow the user to choose the data configuration the user would like to use.

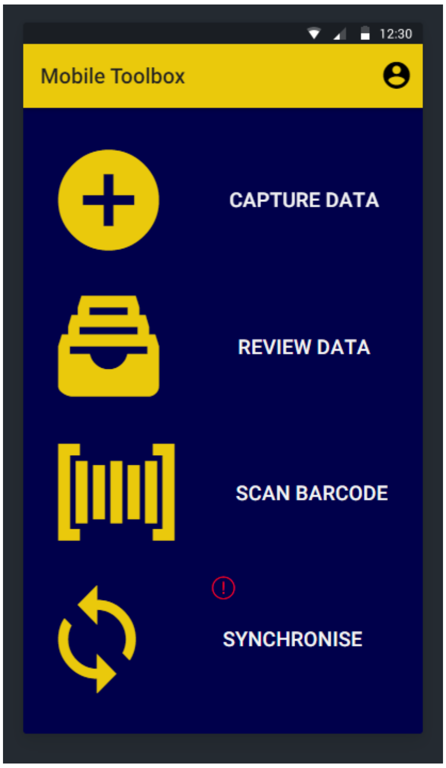


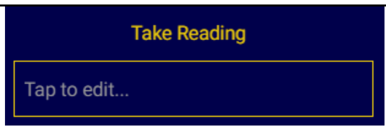
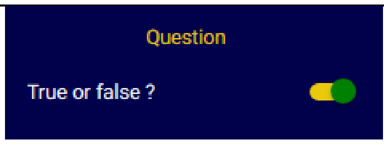
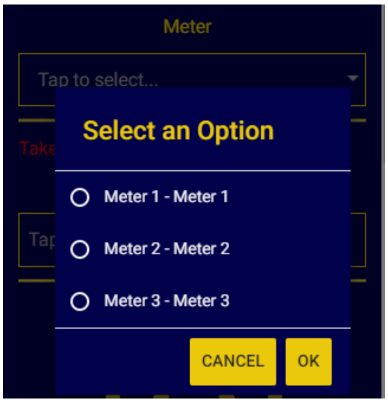
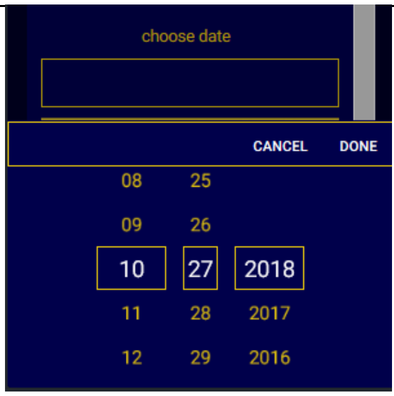
Figure 23: Data consolidation application main screen

2.3.5 Dynamically generated input page

This page will be used to collect data on the system. This page changes based on the inputs received in every configuration. For every data field container in an array, an input is generated. The user interface is developed based on the input type selected for the database. For every set of data collected, the application generates a log container to store the data. For every data field in the configuration, a log value container is developed that can hold the data collected in that data field.

The dynamically generated input page can be generated from two sources: a log container or a configuration container. This allows the program to edit logs even if the configuration used to create the specified record has changed. Finally, the program checks if it can dynamically populate the background used to create logs, such as list options when editing the log. If the application fails, the input is replaced with a text box to ensure data can still be edited. Table 4 below shows a list of inputs available.

Table 4: List of available inputs

Input type	UI	Use
Number or String		*This is just a simple input that will be able to take a string or a number depending on the data type*
Boolean		*The Boolean operator is either in the on or off position. Therefore, this field always starts in the off position.*
Camera		*Allows the user to use the device's camera to take a picture, and the image will be stored in the log*
List selector	 	*These are several pre-defined list options that the user can select from.*
Date picker		*Pressing on the input creates a dialog box allowing the user to pick a date.*

2.3.6 List options

One of the advanced features added to the program is to link fields. This is used to ensure the user inserted the correct information. For example, as depicted in Figure 24, if the user selects meter two, the application will show five options for the user to select and populate the data field.

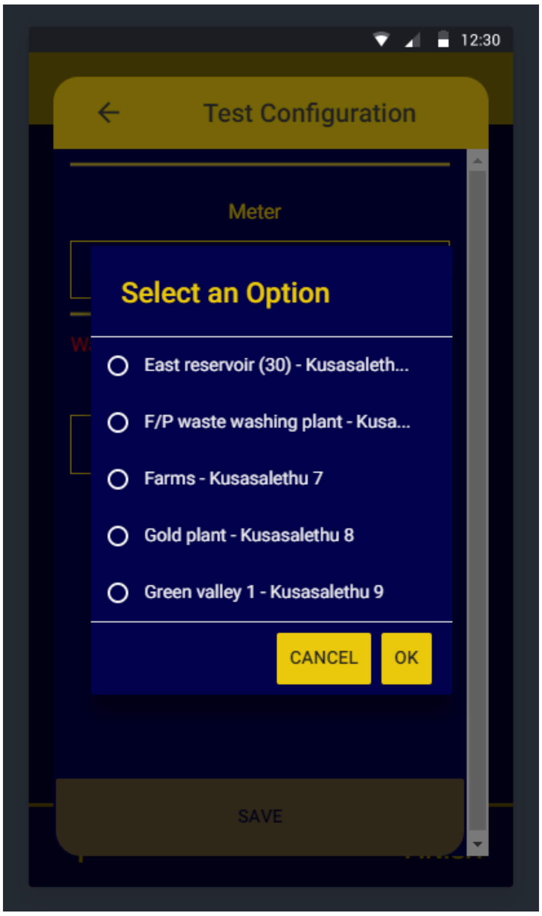


Figure 24: List option field in application meter 2 picked

If the user selects one, as in Figure 25, there are only four options that the application shows the user to choose from. These options can be set up for any field with any value. This feature can make data collection much more straightforward by limiting the answer a user can pick based on previous results. This also is one of the functions that allow the application to validate using other fields which is criterion three in the method.

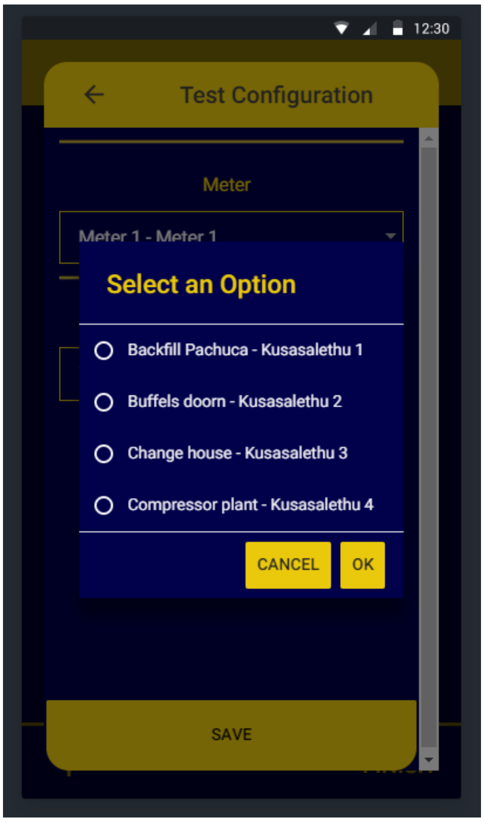


Figure 25: List option field in application meter 1 picked

By selecting the scan barcode option, the application will open the screen shown in Figure 26. Pointing this screen at a barcode will allow the application to read the information on the barcode and check if there has been any information linked to that barcode. A configuration can be opened with the linked information already inserted if the information is connected. This allows users to start collecting data without inserting data that would already be known at that point.

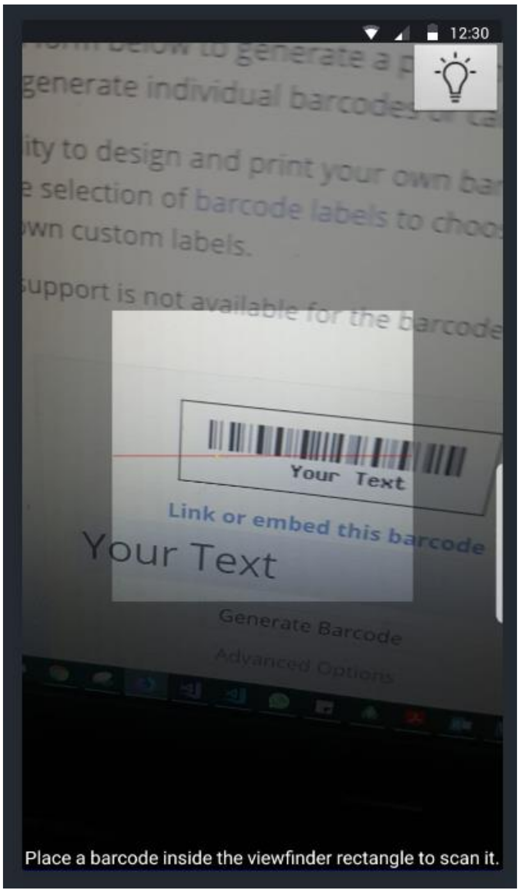


Figure 26: Scanning barcodes in the application

To test the barcode scanner, barcodes were placed on the electricity meter readers, an example of which can be seen in Figure 27. The string saved in the barcode was linked to the location of the electricity meter reader. During the field test, the device scanned the barcode, and the correct configuration opened with the right place already selected. This means that the barcode scanner functioned as designed and programmed.

.

Figure 27: Barcode placed on an electricity meter at mines

After scanning a barcode or selecting the correct configuration, the screen seen in Figure 28 will appear. The user will then fill in this screen. The fields that cannot be skipped are marked as required. The data fields cannot be saved without first filling in these fields. Once all the necessary fields are filled in, the user can click the save button. This will then check the fields and save the data collected.

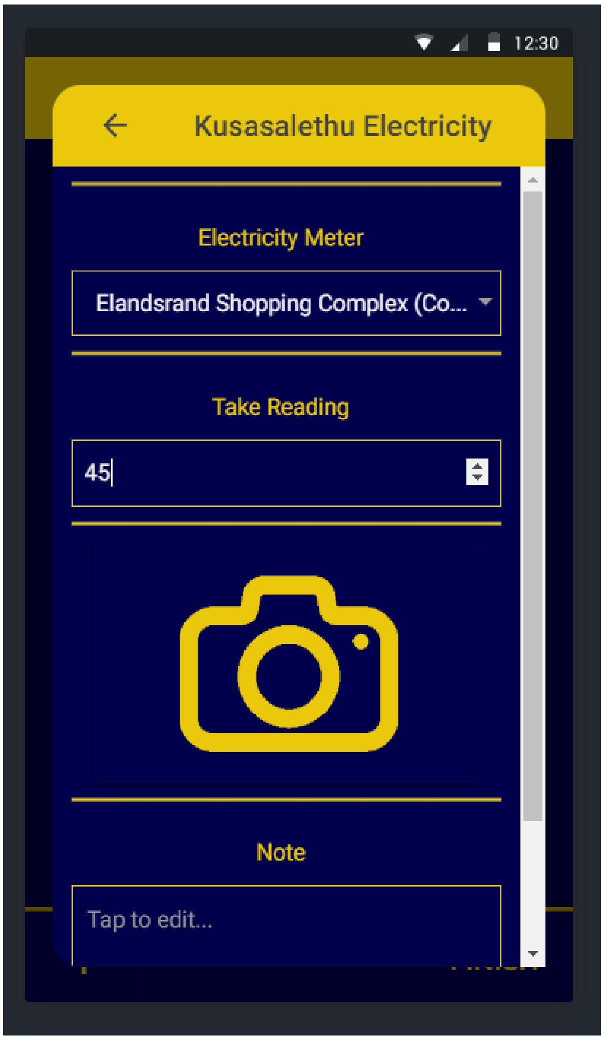


Figure 28: Example of data collection in the application

Users will then go to a location that has an internet connection to synchronise their logs. The next shift worker will take the next device and log in. Having redundant fields that are connected such as a camera connected to a number input, completes criteria three, namely validation using other fields. Being able to log in, open the configuration and collect the data offline completes criteria four, working offline.

2.4 Log retrieval and communication development

2.4.1 Data synchronisation

The fourth step in basic data consolidation is to synchronise the data with an online server. This will allow the data collected by users to be consolidated in one location. Also covered in this section is criterion five, which pertains to encrypted communication with the online server. To use our existing data analysis framework, the application will have to be able to communicate with the online server using Application Program Interface (API). An API is a basic set of instructions that servers and computers share. The device will have to be connected to a communication medium like the internet during synchronisation. To ensure the security and reliability of the data transfer, data should be encrypted and authenticated to ensure the data reaches the server correctly and securely.

The collected data is stored in logs. These logs are sent to the application from the last record synchronised to the present. To request these logs, an encrypted file with the configuration ID, the database and the previous log that was synchronised is sent to the requested log's Uniform Resource Locator (URL) or web address. The request returns an encrypted file containing the logs or an error message. A visual representation of this request can be seen in Figure 29.

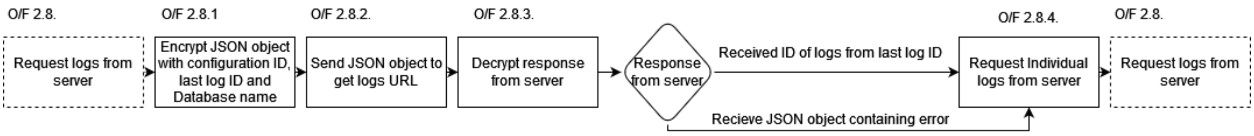


Figure 29: Functional flow - O/F 2.8 request logs from the server

For data collected by the device to be accessible on other devices, the logs are uploaded to the server. The upload process is represented in Figure 30. The new logs are saved and encrypted to a JSON object and then sent to the upload logs' URL. The server then assigns a unique ID to each record and returns this ID either as an encrypted file or as an error message stating why the logs could not be saved.

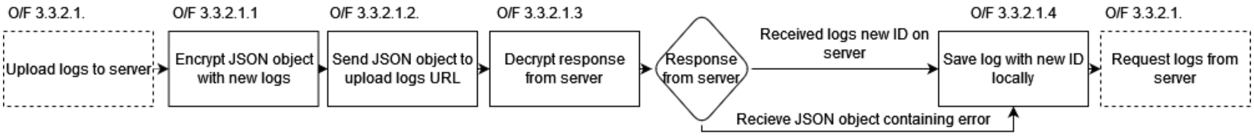


Figure 30: Functional flow - O/F 3.3.2.1 send new logs to the server

Synchronising should also start automatically whenever viable communication with the server is established. This will ensure that data is frequently submitted. In addition, it makes data more readily available and reduces the strain on the system from sending bulk data.

2.4.2 Data fields container

These keys store each data field generated for a selected configuration. The data structure presented in Table 5 is used to generate a dynamic input and storage location for each data field. The system will also allow for basic validation, like required fields and visibility. Finally, the application uses the input type field to generate the correct user interface for the data field.

Table 5: Data field data structure

Key	Storage structure	Use
ID	Unique number identifier	The unique code is used to identify data fields on the application and the server.
PrimaryText	The heading of the data field	This key is used in the UI to inform the user what data will be inserted into this field.
Visible	Boolean set if the data field should be displayed	It can be used to hide data fields from the user.
Required	Boolean set if the data field is allowed to have null values	If the required field is on, a basic level of validation is applied to the data field. The value of this field is type dependent and cannot be left out.
InputType	A string that stores what type of data should be collected	This value is used by the application to generate the data field dynamically. By setting this field, several types of data can be collected based on the chosen value.
Order	A number indicating the order in which the data field should appear	Used to make the configuration collection page order each data field the way the user requires.
List options	An array of list options	If the data field is of the list type, this key stores an array of list options from which the user can pick.

2.4.3 List options container

If a data field presented in Table 5 is specified to have an input type of list selector, an array of list options will be passed along with the data field. This array will allow list options to be generated and selected from the specified data field. The list option container is seen in Table 6.

Table 6: List option data structure

Key	Storage structure	Use
ID	Unique number identifier	Used by the server to identify list options
ParentID	An ID of a parent data field for the list option	One of the advanced options in this system is fields changing depending on other inputs. In addition, this ID allows the application to link to data fields.
PrimaryText	The value displayed to the user	This value is displayed and indicates to the user the option chosen
SupportText	The value saved when options are chosen	Contains the data collected by this data field
Order	Number	Indicates the order in which this list option should appear
Visibility	Boolean	This key indicates if the list option should be visible to the user

2.4.4 Log container

Logs contain the information gathered by a user using the system. The user's location and date information are stored alongside the values to make data trustworthy. Connection to the server will allow logs to be synchronised, so an extra field is required to hold the synchronisation status between the application and the server. The log container used in this project can be seen in Table 7.

Table 7: Log container data structure

Key	Storage structure	Use
ID	Unique number identifier	Used by server and application to identify the log
ConfigurationID	The ID of the configuration linked to the log	This key is used to link the log to the configuration of the record. This allows the program to gather additional information about the log.
UserID	The ID of the user that collected the log	The user ID holds the user accountable for taking logs. This is one part of making logs auditable.
HeaderID	The ID of header-log is only applicable if the log is linked to a multi-log	If this key is not null, this log is stored as a link log of a multi-log. All the logs with matching header IDs are opened when a multi-log is opened.
Synchronised	Boolean that keeps track of whether the log has changed since the last synchronise	If the log changes after downloading, this key is false, indicating that the record should be uploaded again.
Date	Date the log was initially taken	The Date is another key that holds the user accountable and is necessary for audits. This key is also used to group logs by date.
Location	GPS coordinates where the log was initially taken	This is the last key in the log that holds the user accountable. The location can be used alongside the data collected to ensure the validity of the data collected.

Values	An array of value containers	In this array, there is a log value for each data field linked to the configuration in which this log was created.

2.4.5 Log values

For every log, an array of log values is stored. This data structure will be used to keep the actual data collected. In addition, the metadata field seen in Table 8 will be used to store data that is not linked to the data the user wants but which is needed to regenerate the data collected by the user when the log is opened.

Table 8: Log values data structure

Key	Storage structure	Use
ID	Unique number identifier	Used by the server to identify and link values to logs
Value	String	The data saved to the server from this log value
Metadata	String	Extra data can be required to repopulate the log but is not linked to the data field which is used when the record is edited
PrimaryText	The header of the log	Displays above the log input and informs the user what the field requires.
InputType	A string value indicating the type of data collected in the log	This key is used when editing a log. Used by the application to recreate the original configurations with the collected data filled in where applicable.

2.4.6 Log retrieval website

The next requirement of the case study is allowing the user access to the consolidated data. When the data is synchronised, and the user would like to retrieve the data, the back-end service seen in Figure 31 is used. This system will consolidate all the logs that have been collected up to that point and display them to the user. Since each configuration was set up to collect specific datasets, each configuration shown will have the structure set up. The users authenticated with the correct permissions can then select, download, and edit the desired data. All changes made to data after collections are stored as new logs, leaving a completely traceable log to the original data, if needed.

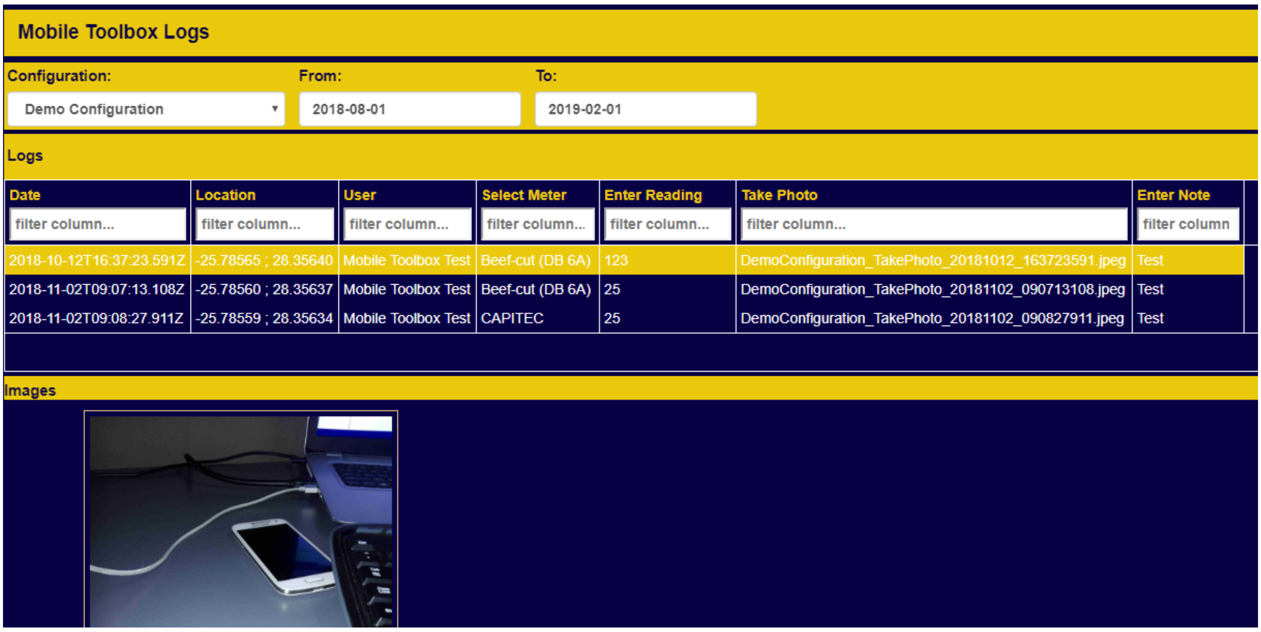


Figure 31: Example of collected data displayed on the server

The data is traceable back to the source with the additional information that is logged such as the user that collected the data, the time the data was collected, and GPS coordinates at the time of data collection, making the data more reliable. This data is also available from the online server and can be accessed at any location with an internet connection. Having the data on a secure server by using encrypted communication satisfies criterion five, an encrypted online server.

2.5 Consolidation system

2.5.1 Summary of system requirements

The mobile interface was designed with ease of use in mind as was investigated in 1.1.6. The system is set up to generate inputs and display them to the user in the order configured. To start data collection, a more straightforward method of barcodes was added. The data can be collected, stored offline, and then synchronised with the online server. Using these methods, all the criteria set out in the methodology chapter are possible; firstly, the steps for consolidation applications:

1. Design forms for data collection in section 2.2.1
2. Give out forms to field devices in section 2.3
3. Collect data with mobile devices in section 2.3
4. Submit data to a centralised database in section 2.4
5. Process data as it is received in section 2.4
6. Provide results and reports in section 2.4

Secondly, the criteria found for the project in Chapter 1 Sections 1.3 and 1.4:

1. Reconfigurable surveys in section 2.2.1
2. Basic input validation in section 2.2.1
3. Validate using other fields in section 2.2.1
4. Work offline in section 2.2.1
5. Encrypted online server in section 2.4
6. Authorise user with secure login offline in section 2.4

2.5.2 System validation

System validation is used to validate that the setup using the application designed in the first part of this chapter can be used to collect the necessary data required for each case study.

The effectiveness of the approach described thus far will be verified by the results of the system in use when the approach is implemented. If the system is of proven monetary value to the users of the system, that will validate the system. A client will ask for their specific data setup requirement, and the data required will be used to create a configuration of the system. The format will be synched to a device running the application, and the application will then complete a unique form and data structure to collect the required data. Operators will use the system to collect data in a harsh industrial environment where they do not have an internet connection. This data will then be synched if the mobile devices are connected to an internet connection when the data collection process is completed. After successfully synching the data, it will be available to engineers. The system will store each unique data configuration and return it to the selected engineer. The project engineer will then use the data in their applications to provide services that would have otherwise not been applicable without the data availability. They will report back any monetary value retrieved from using the data that is consolidated using the system. If the project engineer sends back proof of the economic value, the system can be seen as viable.

2.6 Conclusion

This chapter discussed the mobile architecture suited for data consolidation using the mobile approach in section 2.1. Furthermore, the implementation of the user interfaces and one of the optimal approaches to designing it was shown in sections 2.2 and 2.3. In the same sections, system verification was investigated by discussing the configuration of the individual projects and how the system would allow each configuration to be created.

Next in section 2.4, how the user would use the system to collect the data and develop logs was discussed. How the existing system would be used to integrate the logs with an online server and then to make the data readily available was also discussed in this section. Lastly, section 2.5 presented an overview of the validation method that will be used in this dissertation.

3 RESULTS

3.1 Preamble

This chapter will use the developed method to implement a data consolidation project on two South African mines. These mines were chosen as case studies because of two factors. Firstly, the mines collected paper-based logs of their diesel usage. Therefore, no extra personnel had to be allocated to the study, and the mine could operate normally. The application was used by the operators to collect the data where these operators would normally use paper logs. The aim was to replace the use of the paper-based system with that of the application. Secondly, there were already project engineers assigned to find a project that could potentially have a monetary benefit. This means that the data could be processed, and a suitable project could be identified.

These mines also create a harsh environment for data consolidation since mobile devices will not have an internet connection available while collecting the data. The result of switching to digital collection and data consolidation methods by using the mobile collection systems discussed in previous chapters will be highlighted in this chapter.

3.2 Case study 1: Mine A

Mine A had collected their diesel logs for some time but could not use them for optimising the system since the data had to be digitalised first. However, since the diesel Mine A uses is an operational expense, a diesel rebate was possible if the mine could meet South African Revenue Service (SARS) rebate data demands.

The paper-based logs used to capture data were not traceable and would need to be digitised to give SARS the data needed. SARS also required extra evidence that the data was correct, which was impossible with the paper logs taken. Strict data requirements could be met using a mobile application to collect the data. Time could also be saved by the data not needing to be digitised after collection.

The data already being collected meant that Mine A already had the personnel available. With a prospective way of monetising the data collected, Mine A was a viable candidate for this study. The mine gave the exact data requirements that would be necessary, and that was then translated into a unique configuration for their data needs. The following section shows the setup to create the layout of data collection required for the mine.

3.2.1 Setup of Mine A: Data fields

Applying the method developed in Chapter 2, the first part required for the study is the data the mine needs to collect. This is then translated into data fields: setting up the configuration and linking it to the user completes the first step in the method, namely setting up the system. Setting up a configuration as requested that matches the user's requirements also completes the first criterion, namely reconfigurable surveys. The data fields needed were determined by translating current paper-based logs and adding digital features such as list and pictures where necessary. The data fields that should be configured for this specific case can be seen in Table 9:

Table 9: Setup of diesel usage log for Mine A

Name	*Input Type*	*Use Case*
Quantity Received (Litres)	Number	The number of litres put into the machine and collecting this information is the core of the diesel rebate.
Quantity Received Photo	Camera	Parent Field: Quantity Received (Litres) A photo of the odometer of the petrol station was also used to check the value of quantity and total litres inserted.
Engine Running Hours/Odometer	Number	The number of running hours or km travelled on the machine can be used to ensure that the diesel inserted into the engine has been used in the machine's operation.
Engine Running Photo	Camera	Parent Data field: Engine Running Hours/Odometer A photo of the odometer was also used to confirm running hours.
Fill Up?	List Selector	Choose from a list of why the vehicle stopped at the diesel bay: • Fill Up • Transfer • Top Up
Mine A Location	List Selector	This field separated the diesel pumps in Mine A. 6 locations in Mine A can be used to get diesel.

Mine A vehicle registration / Machine serial number	List Selector	The list of vehicles in this field will mostly be filled in by the barcode being scanned on the vehicle when the data is collected. There is a total of 294 vehicles linked to this data field.
Pump Meter Reading Value	Number	The total litres added at the pump; this field is also used to confirm litres put into the machine and to ensure the next machine value.
Operator Driver Name	Text	Name of the vehicle operator; this field is also used to increase the reliability of the data.
Comments	List Selector	List of possible comments, such as a barcode not working, or the pump having been refilled.
Comments / Other	Text	If there are additional comments to add that are not available on the previous list.
Diesel Bay Operator	List Selector	This field allows the diesel bay operator to choose from a list if anything unexpected happened while collecting the diesel log. The list options are as follows: • Barcode invalid • No barcode • The km/hour gauge is not working • The meter reading gauge is not working • No operator available • Replace meter reading gauge • Testing equipment • Maintenance • This data can then be acted on after the shift if necessary.

The fields related to an input type satisfy the second criterion, namely basic input validation since the application can, by using inputs, ensure the correct data type is collected. The fields marked with parent fields are validated by other fields, either containing identical information but in different formats or filtering other fields with the value inserted; this completes the third criterion, validating with other fields.

3.2.2 Setup of the Mine A barcode

The mine also chose to link various machines running on diesel to barcodes. Consequently, a barcode was created for each vehicle, and attached to the vehicle to be filled with diesel. The operator can scan this barcode, subsequently opening the associated configuration automatically. The opened design will have a data field filled in that was pre-configured.

An example of the data linked to the barcode is as follows:

- Barcode code: 594912
- Configuration: Mine A Diesel Usage Logbook
- Data field: Vehicle Registration Number / Machine Serial Number
- Value: CAM 1122

By filling in this field for the user, not only does the operator not have to navigate the application and open the configuration that logs diesel usage, but the field, in this case, the vehicle registration, is also automatically filled correctly, saving time, and allowing the operator to start filling in other relevant information directly. In Figure 32, an example of the barcodes used on a vehicle underground can be seen. The code that the barcode scanner reads is also written below the barcode. The barcode was printed on metal to ensure the barcode's longevity.

Figure 32: Barcode placed on a vehicle at Mine A

Mine A linked 264 barcodes, each pointing to a specific machine being operated on Mine A. Since this information could be scanned from the barcode, the user possibly saves time by not having to search through the list for the required vehicle number.

3.2.3 Collecting Data on Mine A

The second step is distributing the forms to the users and the third step in the method is users collecting data using the system.

To ensure this is possible, users were trained on using the application on the rugged device. Logging into the application with the user's profile, any logs taken from that device would then be linked back to the user. If the data is linked, that instance of the application no longer requires the server data to function; this also completes the sixth criterion, which is an authorised user with secure login offline. By scanning the barcode on the machine, a configuration would open with the details for the machine already inserted and ensure mistakes were kept to a minimum. The user would then fill in the amount of diesel poured, and the total litres poured in, and take a picture of the odometer of the diesel pump. An example of a photo taken by a user can be seen in Figure 33 below.

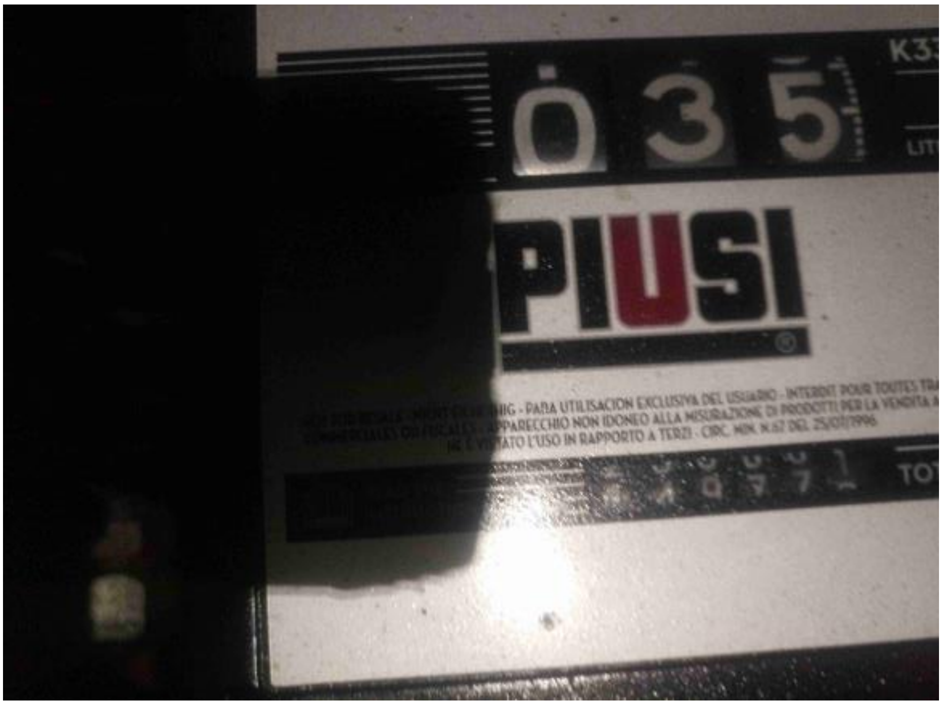


Figure 33: Example of litres added and total

The user will also fill in the engine running hours and take a photo of the odometer on the machine. An example of this odometer can be seen in Figure 34. Having the data in digital format as well as a picture to ensure the data was inserted correctly increases the reliability of the data. All the redundant data collected also helps ensure that the data has not been tampered with, which contributes to the third criterion, validating with other fields.

Figure 34: Examples of engine running hours

After the data was collected and the shift was finished, the operators took the rugged devices to a location above ground with an internet connection. Once there, the devices synchronised the daily logs collected with the online storage systems. This completes the fourth criterion namely the ability to work offline since the devices did not need an internet connection during the data collection process.

3.2.4 Reviewing data collected on Mine A

The fourth step, as referred to in the second chapter, is consolidating the data collected. This is completed once all the devices' data is linked, and data is online. When the data is online, it can be accessed from anywhere with an internet connection. The data can only be accessed by personnel with the correct security clearance. An example of the logs that a user has logged in can be seen in Figure 35.

Vehicle registration number/machine serial number	Cost Code	Choose location	Fill Up?	Engine running hours/odometer	Quantity received (Litres)	Pump Meter Reading Value	Diesel Bay Operator
Filter	Filter	Filter	Filter	Filter	Filter	Filter	Filter
FML 14 - Fermal UVs	26 - R	X/Cut - 8	Top Up -	11814.6	35	149771	Fundiswa -
HD 665 - HD665		T 61 - Tank	Fill Up -	3007.1	30	370180	Nomthand
HD 665 - HD665		T 61 - Tank	Fill Up -	1596.6	93	370273	Nomthand
HD 673 - HD673		T 61 - Tank	Fill Up -	117683	59	370330	Nomthand
FML 15 - Fermal UVs	26 - S	X/Cut - 8	Top Up -	517	36	150123	Kaiser - Mc

Figure 35: Online display of collected data

The fifth step in basic data consolidation is processing the data immediately and the sixth step is to provide output and reports. The data can also be downloaded as a report or as the raw collected data. When the data is downloaded as a report, the redundant information, such as photos, is shown alongside the values scanned. The report was then used to filter out most of the user errors. An example of the report generated can be seen in Figure 36 below. A complete report can be seen in the appendix.

DIESEL USAGE LOGBOOK

(Client name and client premises (if possible) where eligible activity performed to be inserted)

Diesel Used							Balance after Use	Purpose				Eligible Purchases (Litres)
Type of Vehicle / Machine	Opening Balance of fuel in Vehicle / Machine	Vehicle Registration no / Machine Serial no	Opening km / hr meter / engine hr	Closing km / hr meter /Engine hr	Total km/hr metre /engine hr Used	Total quantity of fuel used (Litres)	Unused Balance in Vehicle / Machine	Specific Eligible Activity Performed	Non-Eligible Activity Performed & liters used	When activity performed	Where Activity performed	Total for eligible Litres
Grader	**149**	CAM4016	13926.5	13944	17.5	**268.53**	**131.47**	Scrape and level the road surfaces	N/A	01-Nov-20	pit	**268.53**
Dump Truck	**286**	CAM1327	5305.9	5311.4	5.5	**139.27**	**310.73**	Cargo transport	N/A	01-Nov-20	pit	**139.27**

Figure 36: Mine A diesel usage report

After the data has been checked for errors, the data can be downloaded in the form seen in Table 10. This data is the form the project engineers use to do the required calculation and further check the data quality. This completes the fifth criterion of communication with an encrypted online server.

Table 10: Raw data retrieved from Mine A

Date	2020/11/24 01:00	2020/11/24 01:00	2020/11/24 01:41	2020/11/24 01:57	2020/11/24 02:03
Location	00.00; 00.00	00.00 ; 00.00	00.00; 00.00	00.00; 00.00	00.00; 00.00
User	User 1	User 2	User 1	User 1	User 1
Mine A Machine serial number	CAM 1205 - Dump Truck	CAM 1210 - Diesel Bowser	CAM 1206 - Dump Truck	V 33 - Dump Truck	CAM 1288 - Dump Truck
Mine A Location	Bowser - 5022	Tank 1 - Pump 3	Bowser - 5022	Bowser - 5022	Bowser - 5022
Fill Up?	Top Up - Top Up	Transfer - Transfer	Top Up - Top Up	Top Up - Top Up	Top Up - Top Up
Engine running hours/odo meter	14315	10288,7	14706	7396	6592,7
Engine running photo	20201124_010008 129.jpeg	20201124_010051 538.jpeg	20201124_014125 278.jpeg	20201124_015741071.j peg	20201124_020354 199.jpeg
Quantity received (Litres)	355	8937	228	300	201
Quantity received photo	20201124_010008 129.jpeg	20201124_010051 538.jpeg	20201124_014125 278.jpeg	20201124_015741071.j peg	20201124_020354 199.jpeg
Pump Meter Reading Value	2966555	55867009	2966783	2967083	2967284

3.2.5 Diesel rebate on Mine A

Using the data, the existing system in the mine could be improved. This was done by the project engineer using the data to compile a diesel rebate report. First, the information required by SARS for diesel rebates was compiled into a report. Then, this report was sent to SARS to get a diesel rebate from them. The monthly report summary discussed with the mines can be seen in Table 11. This table shows the total number of litres poured across a range of assets and the redundant fields to get only the eligible litres that could be claimed according to the SARS standards. This processing was done by the project engineers assigned to Mine A.

Table 11: Mine A, an example of a monthly claim summary

November 2020 logbook		
Total purchases (invoiced)	609 069	**Litres**
Total purchases used	608 456	**Litres**
Non-eligible purchases used	224 849	**Litres**
Eligible purchases used	383 607	**Litres**
Claimable	1 071 030	

This project has been ongoing from November 2020 until the present. At the time of writing, the last month's data received is September 2022. The summary of the amount claimed for diesel rebates can be seen in Figure 37.

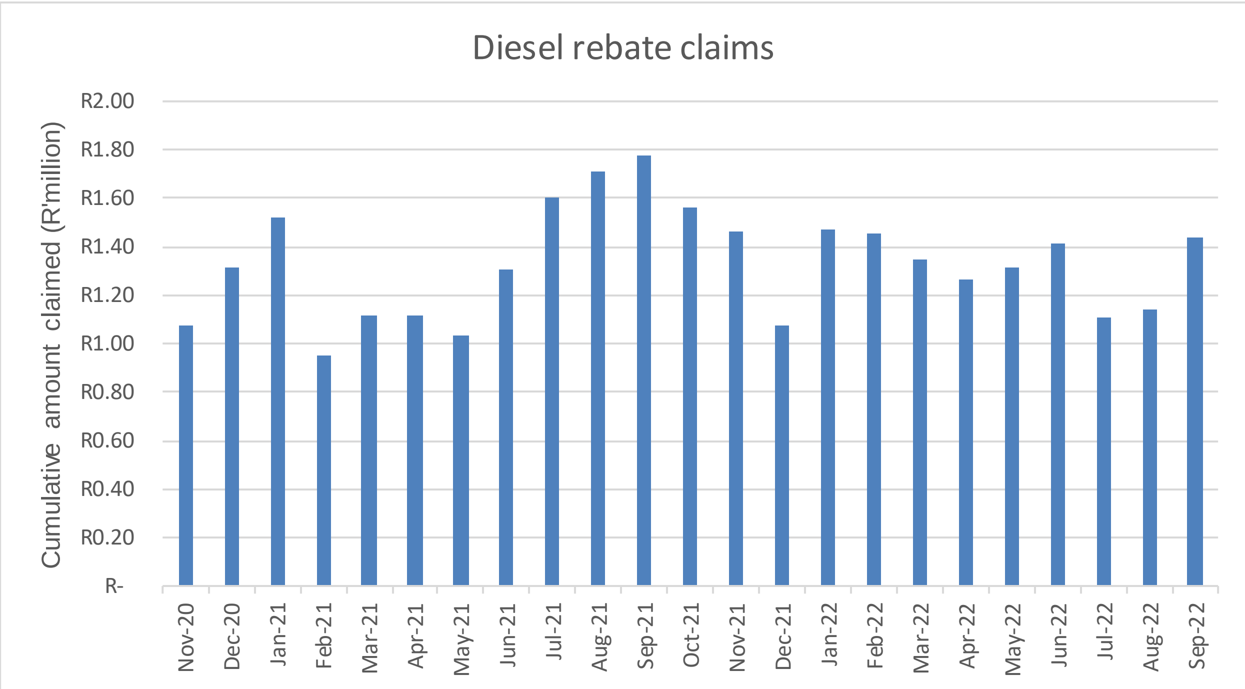


Figure 37: Mine A claimed amounts for diesel rebates

As shown in Figure 37, the mine has an average claim of R 1 328 726.19 per month and a total of R 30 560 702.51. This completes the sixth and last step, receiving feedback on the usefulness of the system. Using the system to consolidate the diesel usage logs collected on Mine A proved to have a monetary benefit. The system was successful in its goals in this regard.

3.3 Case study 2: Mine B

Another mine that met the requirements laid out in Chapter 2 was found. Mine B had been collecting logs for diesel on a paper-based system, so optimising the system using the records was possible for the mine. The diesel collected at Mine B is an operational expense, and if the mine can meet the data demands from SARS, it will qualify for diesel rebates.

Mine B has fewer users to collect data and is a smaller operation than Mine A. Mine B can be used to test the method on smaller industrial systems with less collection overheads. However, several data requirements were not being met at Mine B. Mine B also has different locations and machines to connect to a system and needed some different inputs to be created compared to Mine A, and these can be seen in the next section.

The paper-based logs captured by the mine personnel were not traceable back to the source and would need to be digitised to use the data and send it to SARS. SARS also requires redundancy in the data collected to ensure the quality of the data collected, which also was not possible with the current method of collecting data. The data requirements for a diesel rebate could, however, be met using a mobile application as described in Chapter 2. The mine will also save time with this method since the data will already be in digital format and will not need to be digitised first.

The diesel logs were already being taken at Mine B, which made it fit the requirements in section 2.2.1. Firstly, the personnel needed to collect the data was already available, so no new employee had to be hired. Secondly, project engineers identified a prospective way of monetising the data collected. This meant that Mine B was a viable candidate for this case study.

Mine B then gave data requirements and specifications. These requirements and specifications were then used to create a unique configuration that fit the mine's data needs. In the following section, the setup of Mine B's configuration for their data collection needs will be shown.

3.3.1 Setup of the Mine B data fields

To complete the first criteria of the method of reconfigurable surveys, the system will need to be set up for the user's needs. By taking the current paper-based logs that were being used to collect diesel logs and translating them into digital format each data point needed per diesel log was configured into data fields. The data fields and their purpose can be seen in Table 12:

Table 12: Setup of diesel usage log for Mine B

Name	Input Type	Use Case
Type of vehicle	List selector	This data field allows the user to choose from vehicle types. This list contains the following types: • LDVs (Bakkies) • Fermal UVs • Loader • Trucks • Locos • Agi cars • Forklifts • Graders • Charging up machines • Drill rigs It is choosing one option that filters the vehicle registration data field to only include vehicles of the specified type. This ensures that less time is needed when searching through the list of vehicles.
Vehicle Registration Number / Machine Serial Number	List Selector	This list will be filtered if the type of vehicle field has a value entered or will populate the kind of vehicle if a vehicle is selected from the list. This field can also be filled in by scanning the barcode on the vehicle and is used to identify the vehicle that is getting diesel. 189 vehicles are linked to this field in a list, and each has a barcode connected to the vehicle.
Quantity Received (Litres)	Number	The number of litres put into the machine; collecting this information is the core of the diesel rebate.

Engine Running Photo	Camera	Parent Data field: Engine Running hours/odometer. The odometer on the vehicle was used to confirm engine running hours.
Engine Running Hours / Odometer	Number	The number of running hours or km travelled on the machine can be used to ensure that the diesel added to the machine has been used in the machine's operation.
Quantity Received Photo	Camera	Parent Data field: Quantity Received (Litres) A photo of the odometer of the petrol station was also used to check the value of quantity and total litres inserted.
Fill Up?	List Selector	Choose the purpose of stopping the machine at the diesel bay from the list: • Fill Up • Transfer • Top Up
Choose Location	List Selector	This field separated the diesel pumps in Mine B. 8 locations in Mine B can be used to get diesel.
Pump Meter Reading Value	Number	The total litres added at the pump; this field is also used to confirm litres added to the machine and to ensure the next machine value.
Comments / Other	Text	If there are additional comments to add that are not available on the Comments list selector.
Diesel Bay Operator	List Selector	The logged-in user will fill in a person that added the diesel.
Cost Code	List Selector	The code connected to the vehicle, filtered by the vehicle that is filled in, adds extra information about the amount of diesel the vehicle can hold.

Section	List Selector	Section of mine that the vehicle is being filled with diesel.

This is also the first step in a basic data consolidation application to create data collection forms. The input types and parent data fields connected to the data fields complete criteria two (basic input validation) and criteria three (validating using other fields separately).

3.3.2 Setup of the Mine B barcode

Each vehicle was fitted with a barcode to increase the efficiency of collecting data. These barcodes could then be linked to the machine codes and vehicle types. For example, a barcode field was created to connect each vehicle to its machine code. This barcode field was then linked to the barcode fitted on the vehicle. This allowed the diesel operator to scan a barcode, and the application could automatically open the correct configuration to fill in and place the valid machine number filled with diesel.

An example of the data linked to the barcode is as follows:

- Barcode code: MTB100363
- Configuration: Diesel Usage Logbook
- Data field: Vehicle Registration Number / Machine Serial Number
- Value: HD 666

The application automatically fills this field for the user, contributing to fewer mistakes being incurred, since the user cannot accidentally select the wrong machine. Filling in the field saves time since the user does not need to navigate the application to open the correct configuration.

Since the field is also automatically selected, the user does not need to scroll through the possible vehicles to find the valid vehicle code. The vehicle registration field also has the most available options for this configuration, allowing the user to quickly collect the rest of the data needed for the diesel log.

In Figure 38, an example of the barcodes used underground can be seen. The code that the barcode scanner reads is also written below the barcode. The barcode was printed on metal to ensure the barcode's longevity.

Figure 38: Barcode placed on a vehicle at Mine B

Mine B had 229 barcodes created and linked for the diesel log. Each barcode has a machine code for a machine operated on Mine B. The barcodes attached to the configuration and the vehicles of Mine B will attempt to save time and increase efficiency by cutting down the time that the user would need to fill in logs and find configurations. The following section discusses the method by which the data is collected.

3.3.3 Collecting data on Mine B

The second step is to distribute the forms to the devices and the third step is for users to use the system to collect data. To accomplish this objective, the personnel in Mine B that had not used an application to collect diesel logs received training on using the application on rugged devices that could withstand the underground environment in Mine B.

Once the user is logged into the application, the user's profile is linked to the logs taken while that user is logged in. This is done to help ensure that the data is traceable back to the user. Once the users complete the login and the data is stored on the device, the user no longer requires an internet connection even when logging in to the application again. This completes the sixth criterion to authorise users with secure offline login. By scanning the barcode on the machine that has arrived for diesel, the application would open a diesel log with the machine data available. This saved time and kept mistakes to a minimum for that data field. The personnel would then insert the amount of diesel added to the machine and the total litres on the odometer.

A photo of the diesel bay odometer can be seen in Figure 39. Logging the image ensures the quality of the data since the amount can be verified by checking the photo.

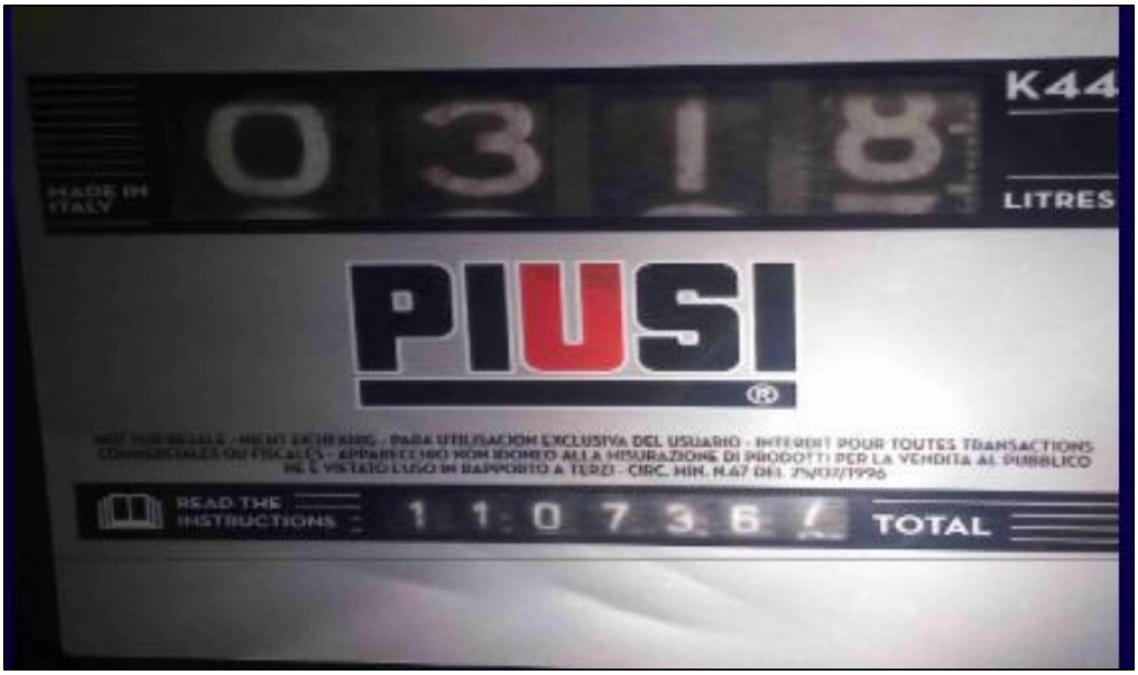


Figure 39: Example of litres added and total litres

The next step for the person filling in the diesel log is to fill in the engine running hours of the machine. A photo is also taken of the odometer of the machine. An example of this odometer can be seen in Figure 40. The redundancy of having the data in digital format along with a picture of the data collected increases the reliability of the data. It can also be used to correct any mistakes later. In addition, the redundant data can be used later to ensure the data was not tampered with. Using other fields to validate completes the third criterion set out in section 2.1.

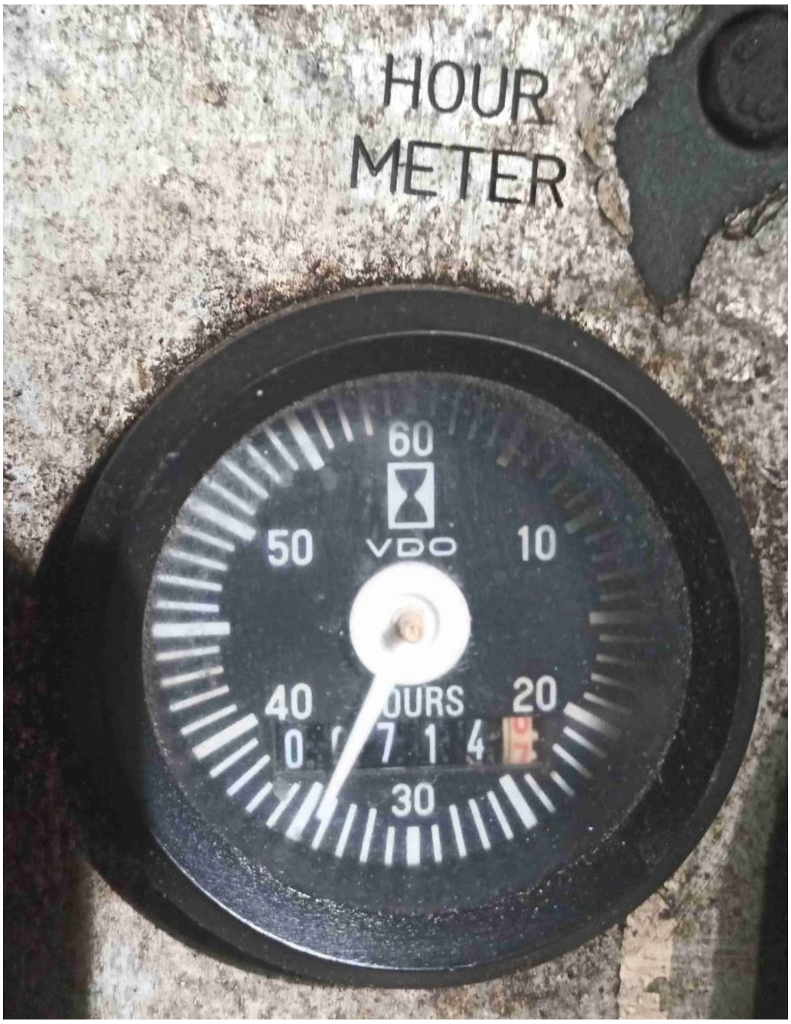


Figure 40: Example of engine running hours

The data is collected offline underground which completes the fourth criterion, which is the ability to work offline. The fourth step in data consolidation is to upload the data to a centralised database. To do this the personnel return the devices at the end of each shift to a location with an internet connection. The devices will then synchronise the data that was collected. Once the data is synchronised, the data is available in the online storage system and can be used by anyone with access and an internet connection.

3.3.4 Reviewing data collected on Mine B

With the data synched from the devices, the fifth step is to process the data immediately and the sixth step is to provide output. The data is now available on the website, completing the consolidation step. The data is only available to users with security clearance to view and or edit the data. This completes the fifth criterion, using an encrypted online server. An example of the logs that a user has logged in can be seen in Figure 41.

Location	User	Vehicle registration number/machine serial number	Cost Code	Choose location	Fill Up?	Engine running hours/odometer	Quantity received (Litres)	Pump Meter Reading Value	Diesel Bay Operator	Requisition no	Section
Filter	Filter	Filter	Filter	Filter	Filter	Filter	Filter	Filter	Filter	Filter	Filter
-27.76445 ; 26	Diesel Logbc	FML 14 - Fermal UVs	26 - R	X/Cut - 8	Top Up -	11814.6	35	149771	Fundiswa -	646224	T portal
00.00 ; 00.00	Diesel Logbc	HD 665 - HD665		T 61 - Tank	Fill Up -	3007.1	30	370180	Nomthandi	2503163	Stores
00.00 ; 00.00	Diesel Logbc	HD 665 - HD665		T 61 - Tank	Fill Up -	1596.6	93	370273	Nomthandi	2503164	Stored
00.00 ; 00.00	Diesel Logbc	HD 673 - HD673		T 61 - Tank	Fill Up -	117683	59	370330	Nomthandi	2501902	Mechanica
00.00 ; 00.00	Diesel Logbc	FML 15 - Fermal UVs	26 - S	X/Cut - 8	Top Up -	517	36	150123	Kaiser - Mc	646227	ventilation

Figure 41: Online display of collected data

This data can also be downloaded in multiple formats, including a report or as raw data. When the users ask for a report, the program compiles the data and the redundant information, such as photos, are displayed alongside the user input or values scanned. The report can then assist in further removing user errors. An example of the report generated can be seen in Figure 42.

Opening Metre Reading		Purchase Receipts			Metre Reading before Disposal	Disposal					Metre Reading After Disposal	Activity
Date	Opening Balance Reading (Litres)	Invoice Number	Date	Litres Received	Opening Balance plus Receipts	Date	Litres Disposed (if losses occurred, also note here)	DISPOSED TO Each vehicle/unit to be indicated separately		Purpose of Disposal (State whether or Eligible or non-eligible use)	Opening balance plus Receipts minus Disposals	Eligible/Non - Eligible
2020-11-06	46 920	n/a	n/a	0	46 920	2020-11-06	157	Service Truck	CAM5023	Transport equipment and parts	46 763	E
2020-11-06	46 763	n/a	n/a	0	46 763	2020-11-06	168	Dump Truck	CAM1198	Cargo transport	46 595	E
2020-11-06	46 595	n/a	n/a	0	46 595	2020-11-06	191	Dump Truck	CAM1202	Cargo transport	46 404	E

Figure 42: Example of a report generated for Mine B

The users being able to get the data in a report or table format completes the fourth requirement for getting the data to the users. The data is also available in a table format, as seen in Table 13. After the data has been checked for errors, the project engineers can use the data in this format to do the required calculation and further improve the data quality.

Table 13: Raw data retrieved from Mine B

Date	2020/11/24 01:00	2020/11/24 01:00	2020/11/24 01:41	2020/11/24 01:57	2020/11/24 02:03
Location	00.00; 00.00	00.00; 00.00	00.00; 00.00	00.00; 00.00	00.00; 00.00
User	User 1	User 2	User 1	User 1	User 1
Mine B Machine serial number	CAM 1205 - Dump Truck	CAM 1210 - Diesel Bowser	CAM 1206 - Dump Truck	V 33 - Dump Truck	CAM 1288 - Dump Truck
Mine B Location	Bowser - 5022	Tank 1 - Pump 3	Bowser - 5022	Bowser - 5022	Bowser - 5022
Fill Up?	Top Up - Top Up	Transfer - Transfer	Top Up - Top Up	Top Up - Top Up	Top Up - Top Up
Engine running hours/odo meter	14315	10288,7	14706	7396	6592,7
Engine running photo	20201124_010008129.jpeg	20201124_010051538.jpeg	20201124_014125278.jpeg	20201124_015741071.jpeg	20201124_020354199.jpeg
Quantity received (Litres)	355	8937	228	300	201
Quantity received photo	20201124_010008129.jpeg	20201124_010051538.jpeg	20201124_014125278.jpeg	20201124_015741071.jpeg	20201124_020354199.jpeg
Pump Meter Reading Value	2966555	55867009	2966783	2967083	2967284

3.3.5 Diesel rebate on Mine B

The data is used to improve the existing system in the mine. The project engineers compiled a report that SARS requires for diesel rebates utilising the information available online. The report was sent to SARS to claim the diesel rebate from the data collected. The diesel rebate can improve the system by decreasing operational costs. A monthly summary of the data collected will be discussed with Mine B.

An example of such a summary can be seen in Table 14. In this table, a project engineer calculated the number of eligible litres to claim, and the total number of litres added. Then the non-eligible litres, according to SARS standards, are shown. This processing was done by the project engineers assigned to Mine B.

Table 14: Mine B example of a monthly claim summary

	Jun-21		
		Quantity [Litres]	
Total purchases (invoiced)		113 788	Litres
Total purchases used		120 554	Litres
Non-Eligible purchases used		54 664	Litres
Eligible purchases used		65 890	Litres
Claimable (@ 3.66 c/l)	R	192 927,37	

The project started on Mine B in May 2021 and has continued until the present. The last month of data collected before the write-up was August 2022. In Figure 43, the diesel rebate successfully claimed can be seen.

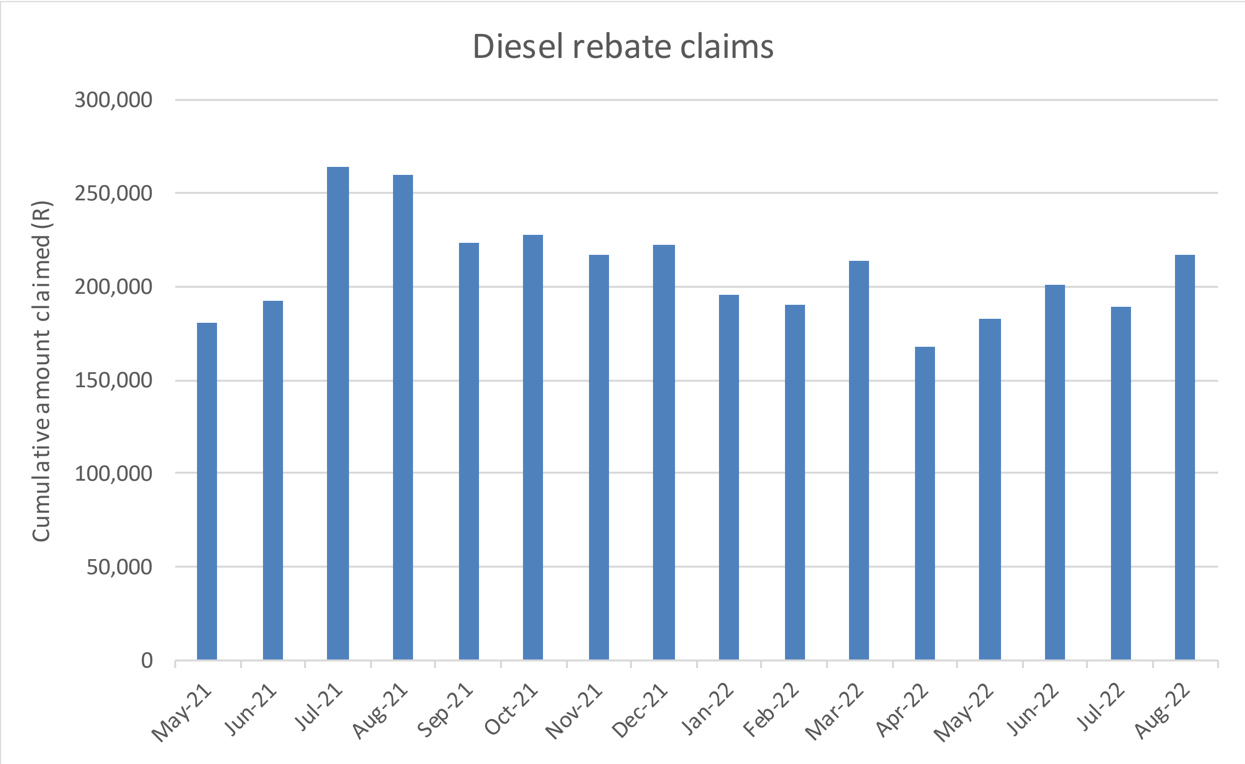


Figure 43: Mine B claimed amounts of diesel rebates

As shown in Figure 43, the mine has an average claim of R 223 528.57 per month and a total claim of R 3 576 457.17. The feedback and monetary value received from the mine complete the sixth objective. Using the system to consolidate the diesel usage logs collected on Mine B proved to have an economic value. Therefore, the system was successful in its goals.

3.4 Comparing diesel rebates to the industrial environment

One question arises: whether these case studies, focusing on diesel data captured which was then used to create diesel rebates, is a fair comparison to the entire industrial environment. The process used in this study can be used in almost any industrial environment for data collection and consolidation because of the inherent setup process. This system was not designed only to be able to collect diesel log data. It was set up to be able to capture data using varying inputs using device hardware such as:

- Input screen
- Camera
- GPS
- Internet connection

Then the system was set up for this specific use case. The fields can have other names, other descriptions, other validation, a different order, or another number of inputs. Since the system could adapt to diesel logs and successfully get the records and synchronise them even when collecting the data in a location without an internet connection, the system should be able to be set up for other applications in the industrial sector.

The limited scope for the case studies still tested the fundamentals of the process that would need to happen in data consolidation, namely:

1. Design forms for data collection
2. Give out forms to field devices
3. Collect data with mobile devices
4. Submit data to a centralised database
5. Process data as it is received
6. Provide results and reports

This system also completed the criteria set out for consolidation in an industrial environment:

1. Reconfigurable surveys
2. Basic input validation
3. Validate using other fields
4. Work offline
5. Encrypted online server
6. Authorise user with secure login offline

The system would also be able to be used for other data collection and consolidation processes in the industrial environment of South Africa, such as water and electricity meter readings, without the need to set up additional reading components. These readings can then be used to detect possible spills or waste.

3.5 Summary

By gathering the data needed to be collected from each mine and using it to create two unique configurations for these mines, the users were able to collect data. The data collected was consolidated and processed. Having the data available enabled the mines to use the diesel logs with the redundant data collected with the system, such as location and pictures, to make a diesel rebate claim to SARS. The data gathered met the SARS requirements for diesel rebates and allowed both mines to get money from diesel rebates.

The application succeeded in the requirements determined from literature 1.3 and 1.4 and listed in section 2.1:

1. Reconfigurable surveys: the application was able to be set up and collect data from two different users in the configuration requested.
2. Basic input validation: the fields inserted would ensure the correct data type was collected for each field selected.
3. Validate using other fields: the application can filter using other fields but also had parent fields that could be linked, to allow users to validate the data inserted by using the data inserted into other fields.
4. Work offline: after the user's first login, no connection to the server was needed until data collection was done.
5. Encrypted online server: communication and storage on the server were encrypted.
6. Authorise user with secure login offline: a method was set up to allow the users to securely log in to the application even when offline.

The case studies had successful claims from SARS of R 1 328 726.19 for Mine A and R 223 528.57 for Mine B on average with a combined total of R1 552 254.76 per month and a total claim of R 34 137 159.69 at the time of writing. This proves that mobile data collection and consolidation of the data has monetary value and can be used to optimise paper-based systems, by making the data more traceable and available. The mines where also able to get diesel rebates from SARS by consolidating the data with the mobile application.

4 CONCLUSION

4.1 Summary

Industries in South Africa are under pressure due to increased operational costs, and to stay viable, industries must become more profitable. Unfortunately, most of the methods found to increase profitability require data.

Data can optimise systems, decrease operational expenditure, and create new projects. Data requirements and restrictions are becoming more advanced, and data digitisation and consolidation can be used to collect data efficiently and reliably.

A system was then designed that could first achieve the fundamentals of the data consolidation process, namely:

1. Design forms for data collection
2. Give out forms to field devices
3. Collect data with mobile devices
4. Submit data to a centralised database
5. Process data as it is received
6. Provide results and reports

Secondly, the system could achieve the criteria set out for consolidation in an industrial environment which was specified for this dissertation in sections 1.3 and 1.4, namely:

1. Reconfigurable surveys
2. Basic input validation
3. Validate using other fields
4. Work offline
5. Encrypted online server
6. Authorise user with secure login offline

The system was configured for these particular use cases to verify that it had met the criteria. As the system could adapt, it could collect data in an industrial setting using diesel logs and successfully retrieve records while offline and synchronise them when connection becomes available. Each need for a consolidation system and a consolidation system in a South African industrial context was satisfied by the system.

The users were able to acquire data by gathering the information required from each mine and using it to produce two different configurations for these mines. The gathered information was analysed and combined. The mines were able to submit a fuel rebate claim to SARS using the diesel logs and redundant data obtained by the system, such as location and photos, because the data was made public. Both mines were able to get money from diesel rebates since the information complied with SARS standards for diesel rebates.

The following conditions, derived from literature 1.3 and 1.4 and specified in section 2.1, were met by the application:

1. Surveys with reconfigurable options: the application might be set up to gather information from two distinct users in the requested configuration.
2. Simple input validation: the fields included would guarantee that the right data type was gathered for each field chosen.
3. Validate using other fields: the program has parent fields that may be linked, which enables users to validate the data entered by utilising the data entered in other fields. The application can filter using other fields.
4. Work offline: after the user's initial login, there was no need to establish a connection with the server until data gathering was complete.
5. An encrypted online server was used for both communication and data storage.
6. Authorise users with safe offline login: a procedure was set up to enable users to sign in securely to the program even when they are not connected to the internet.

The SARS claims for diesel rebates that were successfully submitted and processed by SARS demonstrate that mobile data consolidation and gathering have a monetary value and may be used to improve paper-based systems by making data more traceable and available. To gather data that may be traced, an application was created, and data structures and user interfaces that are dynamically created were used to capture this data. The users were given training on how to utilise this application, and they gathered information in commercial settings, after which the data was synchronised with a secure web server.

After that, the information may be retrieved, processed, and applied to diesel refunds. The project engineers could then use the collected data to create a diesel rebate project. The implementation of this project was tested on two mines. By claiming diesel rebates Mine A saved R 1 328 726.19 and Mine B saved R 223 528.57 which amounts to R1 552 254.76 in total.

4.2 Recommendation for future work

4.2.1 A more comprehensive range of test

The application designed and discussed in the text can collect unique structures of data. Using this feature of the application, a more comprehensive array of data can be collected, such as collecting water or electricity usage logs. This could lead to more monetary value when investigated.

4.2.2 User efficiency

Designing the application in separate ways, such as fewer screen options or blocking the users from completing specific tasks, could affect the efficiency and efficacy with which data is collected. Other designs can be implemented on the users collecting the data to understand which designs work better for users or which methods require less training.

4.2.3 Improved error handling

Using the redundant data captured, such as photos, while the users are busy collecting data rather than being post-processed, could possibly improve error handling. Implementing this change could decrease user errors and increase the reliability of the data captured.

4.2.4 Barcodes for efficiency

More information can be added to the barcodes, like the vehicle's tank size. Such additional metadata can then be used in error handling to ensure the data collected has fewer human errors that need to be filtered and checked. It could also increase the speed and efficiency with which data is captured.

4.2.5 Check the efficiency of vehicles or users

While collecting and digitising information, several monetary gains were discovered that were beyond the scope of this paper. More data points are available such as which vehicle was filled at which pump and the operational hours of that vehicle leading to error detection.

4.2.6 Better control over resources

The collection and digitising of this information allowed the processed data to reveal that some vehicles were receiving more diesel than the tank could store. This leads to further monetary gain since lost diesel could be identified and possibly recovered in the future.

Pattern recognition can be used to detect untrustworthy personnel, vehicles, and logistical and operational opportunities in future iterations of the application. Constant data logging gives more information about the vehicle used, which could lead to better servicing schedules and bring in monetary values since fewer breakdowns would occur.

4.3 Conclusions

Using data consolidation and digitising data can lead to new projects in optimising systems and reducing resource usage. This could help South African industries to stay viable. The project was implemented in the industrial environment of two South African mines by executing the software on data already being taken and using the data gathered to implement new projects.

The feedback received from the implemented project saved the mines money. The mines reported a saving of R 34 137 159.69 in total. With a project that will continue saving money monthly, this shows the monetary value of data consolidation in the industrial environment.

5 BIBLIOGRAPHY

[1] Department of Communications, "South Africa Yearbook 2013/2104," *Mineral Resources,* vol. 6, no. 1, pp. 301-311, 2014.

[2] A. Lane, J. Guzek and W. van Antwerpen, "Tough choices facing the South African mining industry," *Journal of the Southern African Institute of Mining and Metallurgy,* vol. 115, no. 6, pp. 471-479, 2015.

[3] H. Marais, "The economic impact of electricity price increases on various sectors of the South African economy," Deloitte, Page 108, 2012.

[4] H. Groenewald, M. Kleingeld and J. Vosloo, "A performance-centred maintenance strategy for industrial DSM projects," in *Proceedings of the Conference on the Industrial and Commercial Use of Energy*, 2015.

[5] R. Maneschijn, J. C. Vosloo and M. J. Mathews, "Investigating load shift potential through the use of off-gas holders on South African steel plants," in *Conference on the Industrial and Commercial Use of Energy*, Cape Town, pp. 104-111, 2016.

[6] Eskom, "Historical Average Price Increase," 1 4 2022. [Online]. Available: https://www.eskom.co.za/distribution/wp-content/uploads/2022/04/Historical-average-prices-and-increase_v20210903_13h00_no-links_External.xlsx. [Accessed 13 7 2022].

[7] B. Mokoena, P. S. Pouabe, Eboule and J. H. C.Pretorius, "Medium voltage consumers' cost evaluation connected to a micro-grid PV system," *Energy Reports,* vol. 8, no. 10, pp. 235-244, 2022.

[8] J. Blignaut, R. Inglesi-Lotz and J. Weideman, "Sectoral electricity elasticities in South Africa: Before and after the supply crisis of 2008," *South African Journal of Science,* vol. 111, no. 9-10, pp. 1-7, 2015.

[9] Mining Journal, "SA miners facing bigger rise in costs," 18 October 2021. [Online]. Available: https://www.mining-journal.com/politics/news/1419707/sa-miners-facing-bigger-rise-in-costs. [Accessed 22 September 2022].

[10] Minerals Council South Africa, "Minerals source," 2021. [Online]. Available: https://www.mineralscouncil.org.za/industry-news/publications/facts-and-figures/send/17-facts-and-figures/1366-facts-and-figures-2020-pocketbook. [Accessed 10 10 2022].

[11] S. Juneja and A. Anand, "Healthcare 4.0-Digitizing Healthcare Using Big Data for Performance Improvisation," *Journal of Computational and Theoretical Nanoscience,* vol. 17, no. 3, pp. 9-10, 2020.

[12] PR Newswire Association LLC, "Alliance Precast Industries Selects Epicor ERP to Support Digital Transformation," PR Newswire Association LLC, Boston, 2018.

[13] A. Goosen, "A system to quantify industrial data," Masters Dissertation, North West University, Potchesftroom, 2019.

[14] M. Hermann, T. Pentek and B. Otto, "Design Principles for Industrie 4.0 Scenarios," in *49th Hawaii International Conference on System Sciences (HICSS)*, 2016.

[15] W. Eckerson, "Data quality and the bottom line," Renton, WA, Tech. Rep., 2002. [Online]. Available: https://tdwi.org/Home.aspx. [Accessed 2022].

[16] M. Hossan and A. Nadeem, "Towards digitizing the construction industry: state of the art construction 4.0," in *Proceedings of International Structural Engineering and Construction*, Chicago, IL, USA, 2019.

[17] Q. Chen, B. García de Soto and B. T. Adey, "Construction Automation: Research Areas, Industry Concerns and Suggestions for Advancement," *Automation in Construction,* vol. 94, no. 5, pp. 22-38, 2018.

[18] J. Hook, D. R. Geissbauer and J. Vedso, "Industry 4.0: Building the," 2016. [Online]. Available: https://www.pwc.nl/nl/assets/documents/industry-4-0-engineering-and-construction-key-findings.pdf.

[19] D. Reinsel, J. Gantz and J. Rydning, "The Digitization of the World," *Data Age 2025,* vol. 7, no. 2, pp. 3-11, 2018.

[20] T. Streule, N. Miserini, O. Bartlomé, M. Klippel and B. G. de Soto, "Implementation of Scrum in the Construction Industry," *Procedia Engineering,* vol. 1, no. 1, pp. 269-276, 2016.

[21] G. Hunter, A. Bregt, G. Heuvelink, S. De Bruin and K. Virrantaus, "Spatial Data Quality: Problems and Prospects," *Research Trends in Geographic Information Science ,* vol. 44, no. 2, pp. 101-121, 2009.

[22] D. Feinleib, "Big data bootcamp: What managers need to know to profit from the big data," in *1st ed. Berkeley*, CA: Apress, 2014.

[23] J. Manyika, M. Chui, B. Brown, J. Bughin, R. Dobbs, C. Roxburgh and A. H. Byers, "Big data: The next frontier for innovation, competition, and productivity," in *New York*, McKinsey, 2011.

[24] O. Kwon, N. Lee and B. Shin, "Data quality management, data usage experience," *International Journal of Information,* vol. 34, no. 3, pp. 387-394, 2014.

[25] C. J. Bester, "THE INFLUENCE OF TRANSACTIONAL DATA QUALITY ON MONETARY," *Faculty of Management Sciences ETDs,* vol. 1, no. 1, pp. 22-25, 2018.

[26] M. Tomlinson, W. Solomon, Y. Singh, T. Doherty, M. Chopra, P. Ijumba, A. C. Tsai and D. Jackson, "The use of mobile phones as a data collection tool: A report from a household survey in South Africa," *Globalization and Health,* vol. 9, no. 51, pp. 2-9, 2009.

[27] J. Veile, M. Schmidt, J. Muller and K. Voigt, "Relationship follows technology! How Industry 4.0 reshapes future buyer supplier relationships," *Journal of Manufacturing Technology Management,* vol. 32, no. 6, pp. 1245-1266, 2021.

[28] J. Schobel, M. Schickler, R. Pryss, F. Maier and M. Reichert, "Towards Process-driven Mobile Data Collection Applications: Requirements, Challenges, Lessons Learned," in *International Conference on Web Information Systems and Technologies*, WEBIST 2014, 2014.

[29] J. Schobel, M. Schickler, R. Pryss and M. Reichert, "Process-driven data collection with smart mobile devices," *Web Information Systems and Technologies,* vol. 226, no. 3, pp. 347-362, 2014.

[30] Instream, "The Problems of Manual Data," Instream, 16 11 2015. [Online]. Available: https://instream.io/en/. [Accessed 17 10 2022].

[31] D. A. Motsoaledi, "Statement by home affairs minister," Department of home affairs, http://www.dha.gov.za/index.php/statements-speeches/1584-statement-by-home-affairs-minister-dr-aaron-motsoaledi-on-the-recruitment-of-10-000-unemployed-youth-graduates-to-join-the-digitisation-project-of-home-affairs-records, 2022.

[32] S. Kinkade, K. Verclas and M. Org, "Wireless Technology for Social Change: Trends in Mobile Use by NGOs," *Access to Communication Publication Series,* vol. 2, no. 1, pp. 11-23, 2008.

[33] A. Neagu, "Figuring the costs of using paper forms," formotus, 05 July 2013. [Online]. Available: https://www.formotus.com/blog/figuring-the-costs-of-using-paper-forms. [Accessed November 2022].

[34] K. S. More, C. Wolkersdorfer, N. Kang and A. S. Elmaghraby, "Automated measurement systems in mine water management and mine workings – A review of potential methods," *Water Resources and Industry,* vol. 24, no. 1, pp. 100-136, 2020.

[35] K. K. Lwin and Y. Murayama, "Web-based GIS System for Real-time Field Data Collection Using Personal Mobile Phone," *Journal of Geographic Information System,* vol. 3, no. 4, pp. 382-389, 2011.

[36] J. Park, "Evaluating a mobile data-collection system for production information in SMEs," *Computers in Industry,* vol. 68, no. 6, pp. 53-64, 2015.

[37] H. Pence, "What is Big Data and Why is it Important?," *Journal of Educational Technology Systems,* vol. 43, no. 2, pp. 159-171, 2014.

[38] J. Ishwarappa and J. Anuradha, "A brief introduction on big data 5Vs characteristics and hadoop technology," in *International Conference on Computer, Communication and Convergence, ICCC 2015*, NagarBhubaneshwar, 2015.

[39] M. Trifu and M. Ivan, "Big Data: present and future," *Database Systems Journal,* vol. 5, no. 1, pp. 32-41, 2014.

[40] M. Mohammadpoor and F. Torabi, "Big Data analytics in oil and gas industry: An emerging trend," *Petroleum,* vol. 6, no. 4, pp. 321-328, 2020.

[41] M. S. Sumbal, E. Tsui and E. W. See-to, "Interrelationship between big data and knowledge management: an exploratory study in the oil and gas sector.," *Journal of Knowledge Management,* vol. 21, no. 1, pp. 180-196, 2017.

[42] D. Khvostichenko and S. M., "Effect of fracturing chemicals on well productivity: Avoiding pitfalls in big data analysis.," in *SPE International Conference and Exhibition on Formation Damage Control*, OnePetro, 2018.

[43] T. Nguyen, G. Gosine and P. Warrian, "A Systematic Review of Big Data Analytics for Oil and Gas Industry 4.0," *IEEE Access,* vol. 8, no. 1, pp. 61183-61201, 2020.

[44] S. Martinotti, J. Nolten and J. A. Steinsbo, "McKinseyAndCompany," August 2014. [Online]. Available: https://www.mckinsey.com/~/media/McKinsey/Industries/Oil%20and%20Gas/Our%20Insights/Digitizing%20oil%20and%20gas%20production/Digitizing%20oil%20and%20gas%20production.pdf. [Accessed 11 October 2020].

[45] A. A. Costa and A. Grilo, "BIM-Based E-Procurement: An Innovative Approach to Construction E-Procurement," *The Scientific World Journal,* vol. 2015, no. 1, pp. 1-15, 2015.

[46] D. K. Agrawal, "Application of SCADA system in Steel Industries," *Journal of Scientific and Research Publications,* vol. 5, no. 6, pp. 407-409, 2015.

[47] H. Krasner, "Ensuring e-business success by learning from ERP failures," *IT Professional,* vol. 2, no. 1, pp. 22-27, 2000.

[48] M. Prasad and A. J. Gaikwad, "Human error probability estimation by coupling simulator data and deterministic analysis," *Progress in Nuclear Energy,* vol. 81, no. 8, pp. 22-29, 2015.

[49] S. Mohanty, M. Jagadeesh and H. Srivatsa, "Big data imperatives: Enterprise big data," in *BI implementations and analytics*, CA: Apress, 2013.

[50] H. Adkins, B. Beye, P. Blankinship, P. Lewandowski, A. Oprea and A. StubbleField, "The Intersection of security and reliability," in *Best practices for designing, implementing and maintaining systems*, Sebastopol, O`Reilly media, 2020, pp. 3-13.

[51] L. Cui, Y. Qu, G. Xie, D. Zeng, R. Li, S. Shen and S. Yu, "Security and Privacy-Enhanced Federated Learning for Anomaly Detection in IoT Infrastructures," *IEEE Transactions on Industrial Informatics,* vol. 18, no. 5, pp. 3492-3500, 2022.

[52] R. Alleyne, "Welcome to the information age – 174 newspapers a day.," The Telegraph, 11 February 2018. [Online]. Available: https://www.telegraph.co.uk/news/science/science-news/8316534/Welcome-to-the-information-age-174-newspapers-a-day.html. [Accessed 5 November 2022].

[53] H. Weinreich, H. Obendorf, E. Herder and M. Mayer, "Not Quite the Average: An Empirical Study of Web Use," *ACM Transactions on the Web,* vol. 2, no. 1, pp. 1-31, 2008.

[54] hiv.gov, "HIV.gov," South African Goverment, 16 5 2015. [Online]. Available: https://www.hiv.gov/blog/using-infographics-hiv-communication. [Accessed 6 6 2022].

[55] R. Bohn and J. Short, "Measuring Consumer Information," *International Journal of Communication,* vol. 6, no. 1, pp. 980-1000, 2012.

[56] S. Rinaldi, A. Gandhi and N. Selviandro, "Usability Evaluation and Recommendation of User Interface Design for e-HAC Application by Using User-Centered Design Method," in *2022 24th International Conference on Advanced Communication Technology (ICACT)*, 2022, pp. 483-490.

[57] C. L. Meinert, "Clinical trials : design, conduct, and analysis," *Monographs in Epidemiology and Biostatistics,* vol. 7, no. 3, pp. 93-113, 2012.

[58] L. M. Friedman, C. D. Furberg, D. L. DeMets, D. M. Reboussin and C. B. Granger, "Data Collection and Quality Control. In Fundamentals of Clinical Trials," in *Springer International Publishing*, Cham, 2015.

[59] M. Zorzetti, I. Signoretti, L. Salerno, S. Marczak and R. Bastos, "Improving agile software development using user-centered design and lean startup," *Information and Software Technology,* vol. 14, no. 1, p. 106718, 2022.

[60] D. René and C. Leroy, "Environmental Profile Report for the aluminium Refining industry," *EUROPEAN ALUMINIUM,* vol. 6, no. 1, pp. 2017-2019, 2021.

[61] V. Chauhan, D. Beaton, N. Hamada, R. Wilkins, J. Burtt, J. Leblanc, D. Cool, J. Garnier-Laplace, D. Laurier, Y. Le, Y. Yamada and K. E. Tollefsen, "Adverse outcome pathway: a path toward better data consolidation and global co-ordination of radiation research," *International Journal of Radiation Biology,* vol. 98, no. 12, pp. 1694-1703, 2022.

[62] Google, "Google Forms," Google, 1 3 2011. [Online]. Available: https://www.google.com/forms/about/. [Accessed 15 11 2022].

[63] I. Berzina and I. Lauberte, "Tactical Model for Constructing a Prototype of Automatized Assessment of Tourism Economic Impact," in *RESEARCH FOR RURAL DEVELOPMENT 2019*, Jelgava, 2019.

[64] Momentive, "SurveyMonkey," Momentive, 24 July 2011. [Online]. Available: https://www.surveymonkey.com/. [Accessed 15 November 2022].

[65] Microsoft, "Microsoft Forms," Microsoft, 6 7 2017. [Online]. Available: https://www.microsoft.com/en-us/microsoft-365/online-surveys-polls-quizzes. [Accessed 15 11 2022].

[66] V. Paul, K. Iijima, M. Driver, J. Wong and Y. Natis, "Magic quadrant for enterprise low-code application platforms," *Gartner Research,* vol. 19, no. 4, pp. 1-62, 2021.

[67] SurveyLegend, "SurveyLegend," SurveyLegend, 10 10 2016. [Online]. Available: https://www.surveylegend.com/. [Accessed 06 09 2022].

[68] DoForms, "DoForms," DoForms, 10 03 2018. [Online]. Available: https://www.doforms.com/mobile-forms-professional/. [Accessed 17 10 2022].

[69] G. Asiki, R. Newton, L. Kibirige, A. Kamali, L. Marions and L. Smedman, "Feasibility of using smartphones by village health workers for pregnancy registration and effectiveness of mobile phone text messages on reduction of homebirths in rural Uganda," *Plos one,* vol. 13, no. 6, pp. 10-19, 2018.

[70] I. Prinsloo, "Development and implementation of an advanced mobile data collection system," in *North West University, Doctoral Thesis*, Potchefstroom, 2016.

[71] FlowFinity, "FlowFinity," FlowFinity, 10 03 2018. [Online]. Available: https://www.flowfinity.com. [Accessed 18 10 2022].

[72] J. Gardner and S. Ostrowski, "Geographic Response Strategies on Canada's West Coast," in *International Oil Spill Conference Proceedings*, Virtual Conference, 2021.

[73] Momentive Inc, "Wufoo," Momentive Inc, 6 10 2014. [Online]. Available: https://www.wufoo.com/. [Accessed 15 11 2022].

[74] Formitize, "Formitize," Formitize, 10 03 2018. [Online]. Available: http://formitize.com/. [Accessed 22 10 2021].

[75] S. Maddox, D. M. Y. Read, H. E. Dalton, D. A. Perkins and N. N. Powell, "Developing a mobile data collection tool to manage a dispersed mental health workforce," *Rural and Remote Health,* vol. 20, no. 1, pp. 5616-5650, 2020.

[76] ClickUp, "ClickUp," ClickUp, 15 1 2011. [Online]. Available: https://clickup.com/. [Accessed 15 11 2022].

[77] E. Johnson, R. Morwane, S. Dada, G. Pretorius and M. Lotriet, "Adult Learners' Perspectives on Their Engagement in a Hybrid Learning Postgraduate Programme," *The Journal of Continuing Higher Education,* vol. 66, no. 2, pp. 88-105, 2018.

[78] ProntoForms, "ProntoForms," ProntoForm, 10 03 2018. [Online]. Available: https://support.prontoforms.com/. [Accessed 28 6 2022].

[79] I. M. Prinsloo, J. N. du Plessis and H. G. Brand, "A data consolidation platform for a web-based energy information system," in *International Conference on the Industrial and Commercial Use of Energy (ICUE)*, Cape Town, 2015.

[80] S. Eglen, "A Quick Guide to Teaching R Programming to Computational Biology Student," *PLoS Comput Biol,* vol. 5, no. 8, p. 1371, 2009.

[81] P. Warden, Big data glossary, Sebastopol: O'Reilly Media, Inc, 2011.

[82] Western Cape Department of Health, "Western Cape Goverment," 2014. [Online]. Available: https://www.westerncape.gov.za/assets/departments/health/healthcare2030.pdf. [Accessed 2022].

[83] A. Boulle, A. Heekes, N. Tiffin, M. Smith, T. Mutemaringa, N. Zinyakatira, F. Phelanyane, C. Pienaar, K. Buddiga, E. Coetzee, R. v. Rooyen, R. Dyers, N. Fredericks, A. Loff, L. Shand, M. Moodley, I. d. Vega and K. Vallabhjee, "Data Centre Profile: The Provincial Health Data Centre of the Western Cape Province, South Africa," *International journal of population data science,* vol. 4, no. 2, p. 1143, 2019.

[84] M. Ghazi and G. Durgaprasad, "Hadoop, MapReduce and HDFS: a developers perspective," *Procedia Computer Science,* vol. 48, no. 1, pp. 45-50, 2015.

[85] A4ACADEMIC, "A4ACADEMIC," 2 6 2017. [Online]. Available: http:a4academics.comtutorials83-hadoop840-map-reduce-architecture. [Accessed 20 09 2022].

[86] A. Haug, F. Zachariassen and D. v. Liempd, "The costs of poor data quality," *Journal of Industrial Engineering and Management,* pp. vol. 4, no. 2, pp. 168-193, 2011.

6 ANNEXURE

1.1 Usage diagrams

1.1.1 Overview

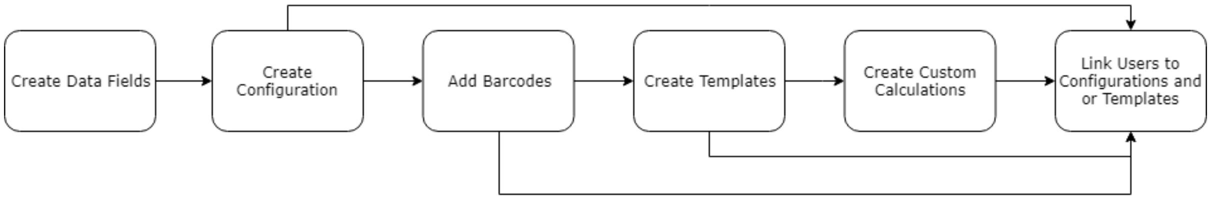


Figure A1: Creation Overview

The diagram shows the high-level steps in creating a Mobile Toolbox **Configuration**. The minimum steps required to create a **configuration** are creating **Data Fields**, creating the **Configuration,** and linking **users** to the created **Configuration**. Adding **Barcodes**, creating **Templates,** or **Custom Calculations** are all *optional*. Ensure that the steps are performed in the order specified.

1.1.2 Data Field Management

This section details the steps in creating **Data Fields**, **List Options**, **Validations** and **Support Items**.

- A **Data Field** represents a value that is collected.
- **List Options** are associated with listSelector (Data Field Input Type). Data Fields
- **Validations** can be linked to Data Fields to ensure that entered data is correct.
- **Support Items** allow for the configuration of contents displayed on log cards.

1.1.3 Adding Data Fields, List Options, Validations and Support Items

This section shows the functionality available on the **Data Field Management** menu.

1. Ensure that the correct Data Field **Input Types** are available within the Input Types sub-tab:
    1. These input types should match the available application input types.
    2. There should not be a need to add these if the selected Mobile Toolbox group has already been initialised.
    3. Input types can be added using a CSV file (Download a template using the Download Template grid button)
    4. The image below shows the possible **Input Types** available at the creation of this document.

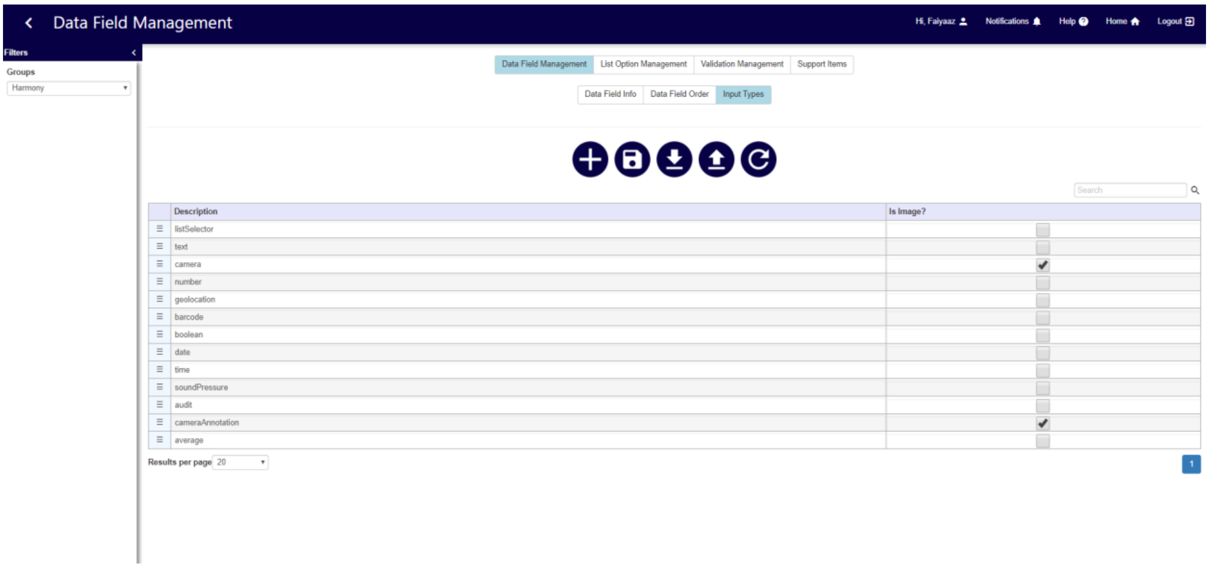


Figure A2: Input Types

2. Add **Data Fields** utilising the **Data Field Info** sub-tab:
    1. Select **Input Types** created using the **Input Types** sub-tab.
    2. **Parent Data Field** refers to a previously created or saved listSelector **Data Field** that serves as the parent of a newly created listSelector Data Field (See List Option Creation for more information)
    3. Data Field **Primary Texts** should not contain any special characters.
    4. Data Fields can be added using a CSV file (Download a template using the Download Template grid button)
    5. The image below shows the *Data Field Info Grid*

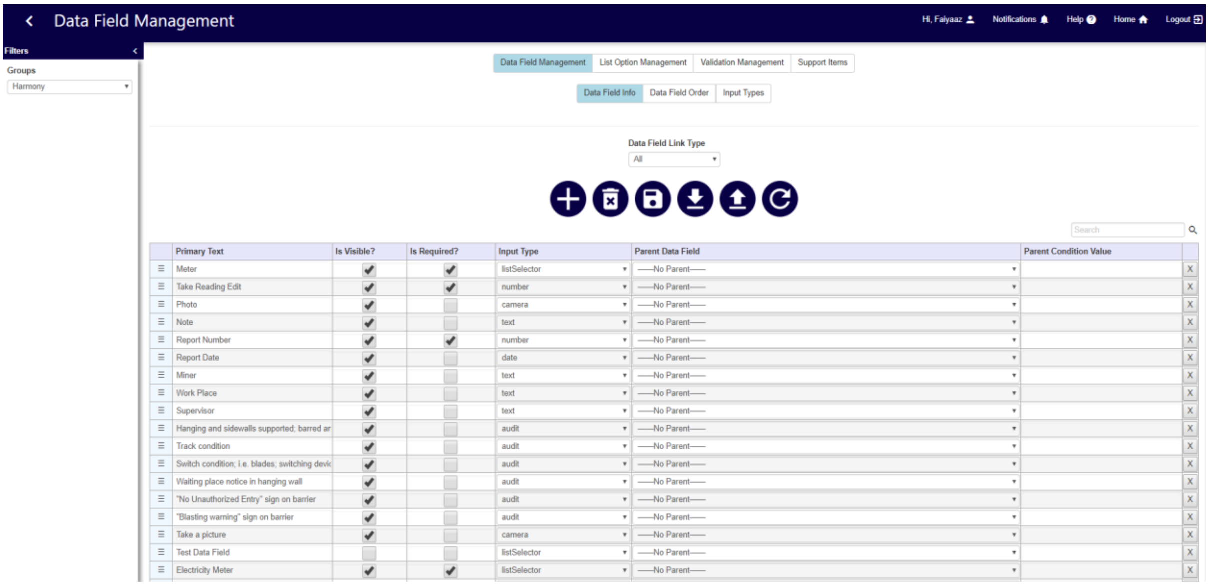


Figure A3: Data Field Info Grid

3. Add Data Field **List Options** using the **List Option Management** Tab
    1. Add **List Options** using the **List Option** sub-tab.
    2. **Parent List Option** refers to a previously created or saved **List Option** linked to an existing **Parent Data Field**

3. List Options can be added using a CSV file (Download a template using the Download Template grid button)
4. The image below shows the *List Option Grid*

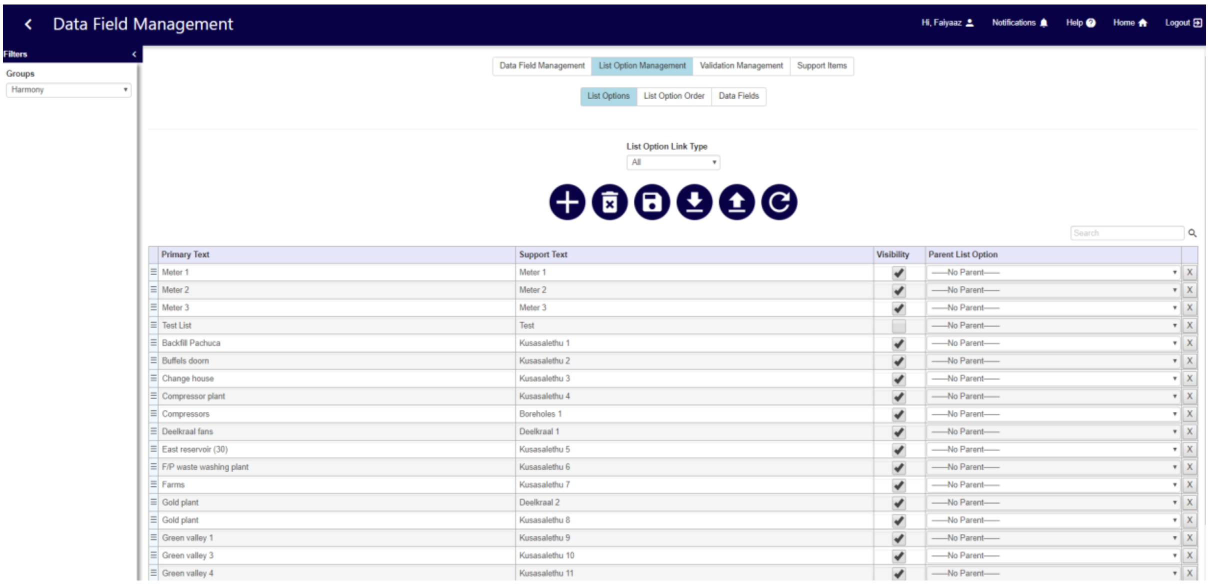


Figure A4: List Option Grid

4. Link **List Options** to **listSelector** Data Fields using the **Data Fields** sub-tab:
   1. Select a **listSelector** Data Field using the **Data Fields** dropdown (Below the Groups dropdown)
   2. Link required **List Options** using the linking table.
   3. **Order List Options** *after* they have been linked to Data Fields using the **Order** sub-tab.
   4. The image below shows the *List Option Link Data Field Linking Table*

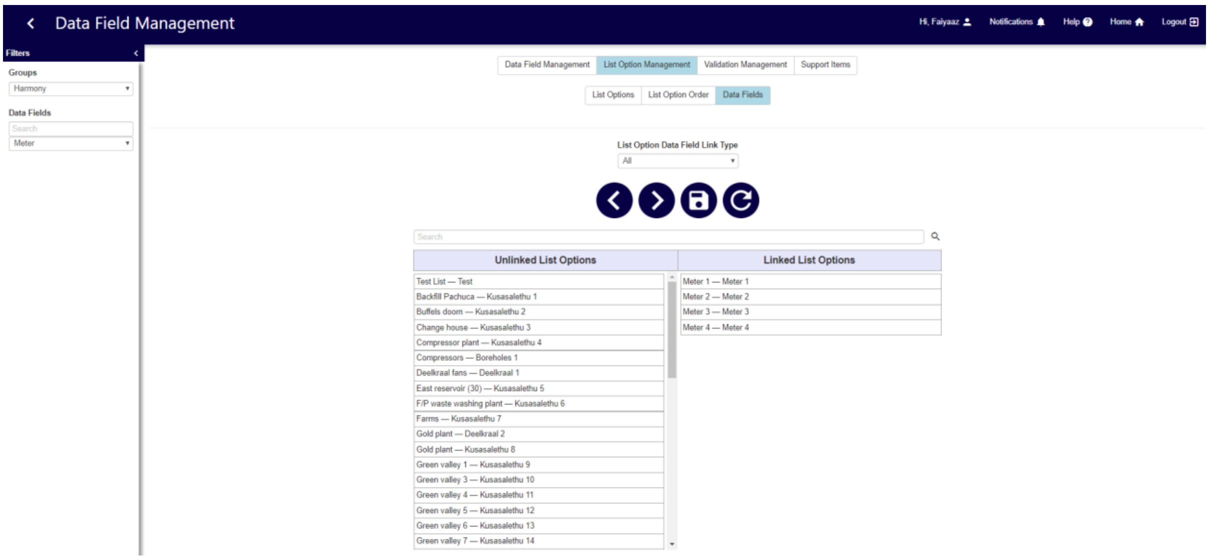


Figure A5: List Option Link Data Field Linking Table

5. Add **Validations** using the **Validation Management** tab:
   1. **Validation Types** are added using the grid within the **Validation Types** sub-tab.
      1. Types should only be added if they are available within the application.

        2. See **Input Types** documentation for details and usage of available types.
        3. The image below shows the *Validation Types' grid.*
    2. **Validation Info's** added using the **Validation Info** sub-tab:
        1. **Validation Info** is used to specify the validation parameters and specific validations for *Data Fields*
        2. A **Validation Info** is linked to a *Data Field*
        3. A **Validation Info** *parameter* can be either a *Value* or be linked to another *Data Field* (The *Data Field* must be linked to the same *Configuration* as the *Data Field* the **Validation Info** will be linked to)
    3. **Validation Info's** linked to Data Fields using the **Data Fields** sub-tab:

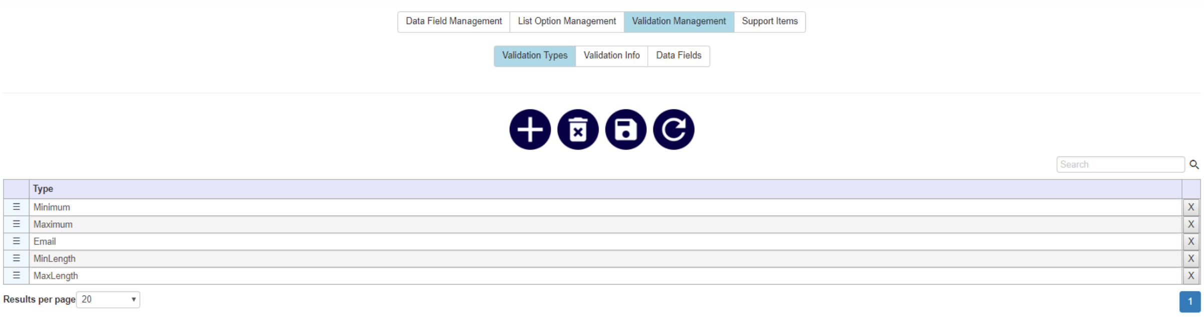


Figure A6: Validation Types grid

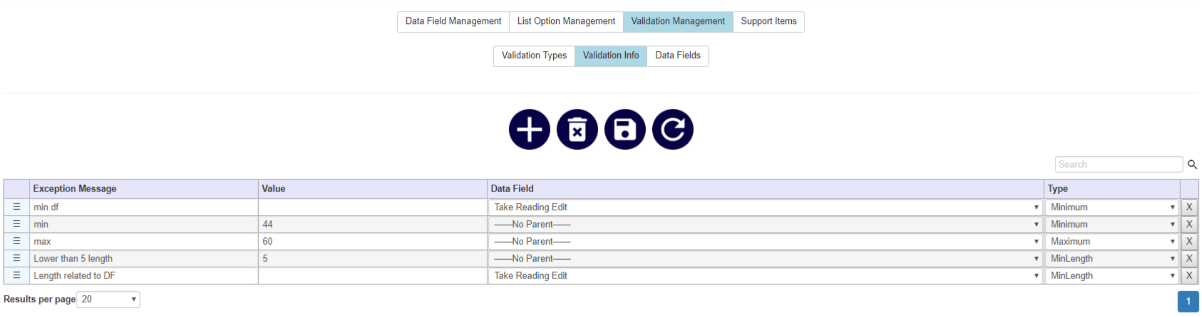


Figure A7: Validation Info grid

6. Add **Support Items** using the **Support Items** tab:
    1. **Image Data Field** must be a Data Field with a camera Input Type
    2. **Historic Identifier** is the primary Data Field that is displayed on a *Log Card*
    3. **Support Text** is the secondary text displayed on a *Log Card*
    4. **Support Items** can be added using a CSV file (Download a template using the Download Template grid button)
    5. The image below shows an example of a *Log Card*

Figure A8: Example Log Card

1.1.4 Configuration Creation

This section details the steps in creating **Configurations** for the Mobile Toolbox application.

- A **Configuration** dictates what data should be captured.
- *Parent* and *Child* allow for two-dimensional data collection.
- An *Orphan* **Configuration** allows for one-dimensional data collection.

1. Ensure that **Data Fields** and **Support Items** have been created.
2. *Optional:* Ensure that **Barcodes** and **Custom Calculations** have been created
3. Add a **Configuration** using the grid:
    1. Ensure that no special characters are used in the Configuration **Heading**
    2. *Parent* and *Orphan* **Configurations** are **Selectable Configurations**
    3. **Create more than one log at a time.** is only relevant for *Orphan* **Configurations**
    4. Child **Configurations** should not be **Selectable.**
    5. When saving a **Configuration**, an **ASTIR JSON TAG** matching the **Configuration** is created.
    6. **Configurations** can be added using a CSV file (Download a template using the Download Template grid button)
    7. The image below shows the *Configurations Grid*
4. Whenever a new **Configuration** is *created*, a corresponding JSON tag is made with the following naming convention: [MobileToolbox] "Configuration Heading" — "Group Name."
    1. Tags can then be linked to reports, and data will be available for use like regular tags.

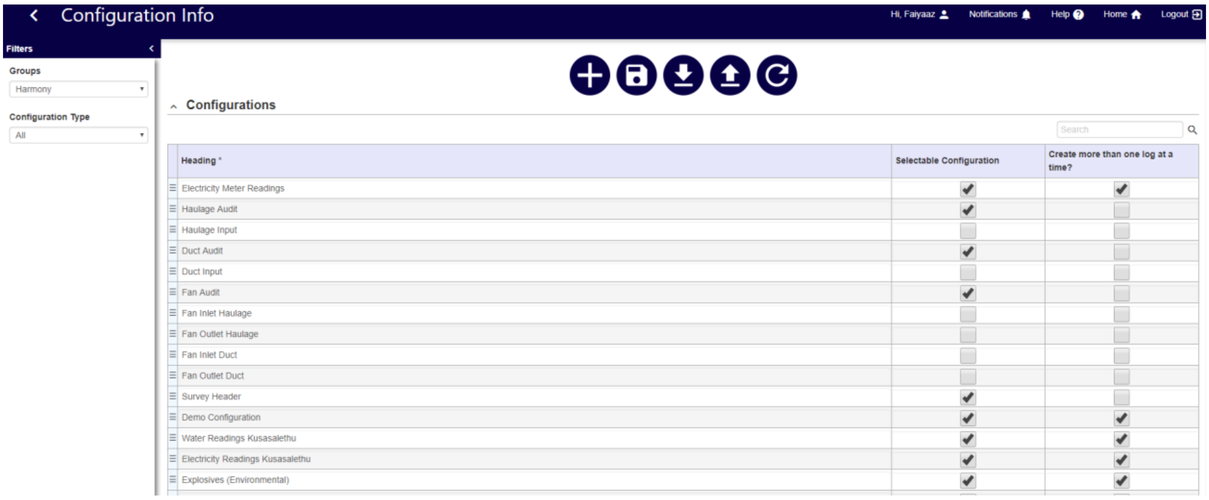


Figure A9: Configurations Grid

4. Save the new **Configuration** and click the Configuration that **Data Fields**, **Support Items** and **Custom Calculations** should be added for
5. The **Child Configurations** tab should be used when creating two-dimensional configurations (*parent* and *child* relationships)
6. The **Data Fields** tab is used to link **Data Fields** to the clicked **Configuration.**
    1. *Optional:* **Order** Data Fields after they have been linked to **Configurations** using the **Order** sub-tab (found within the **Data Field Management** menu)
7. The **Support Items** tab is used to link **Configuration Support Items**
    1. All configurations should have a **Support Item attached** (If they do not, they **cannot** be used on the application)
8. *Optional:* Link **Custom Calculations** to the Configuration using the **Custom Calculations** Tab
9. The image below shows the tabs available when selecting (clicking) a Configuration *Tabs*

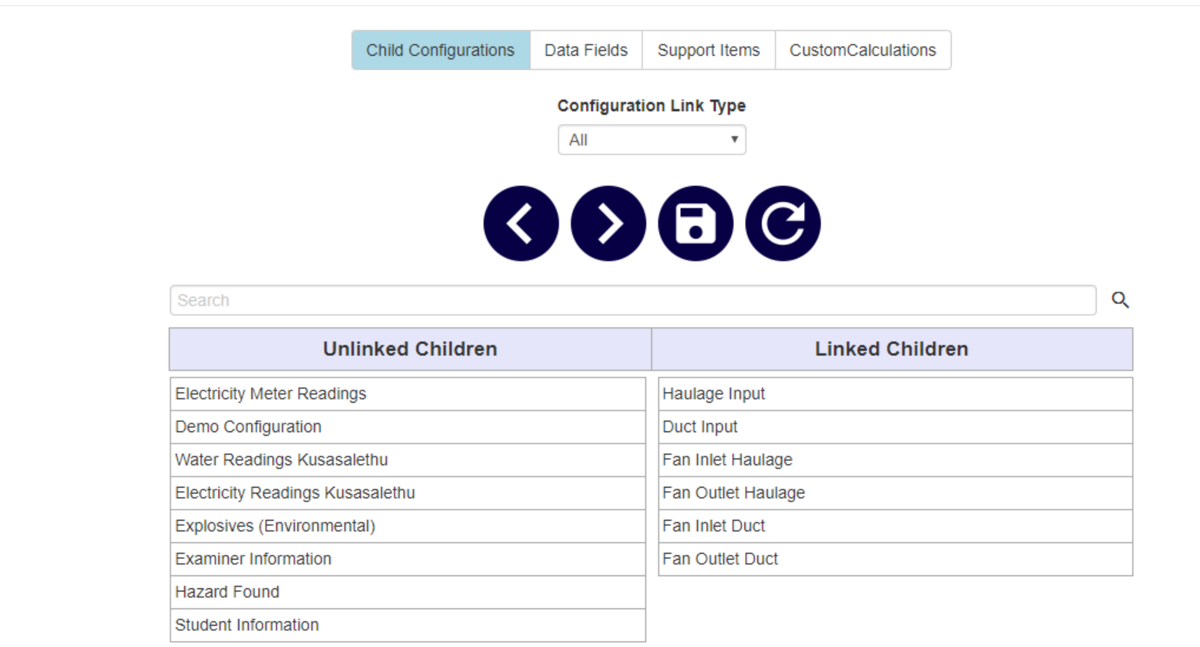


Figure A10: Configuration Tabs

1.1.5 Barcode Management

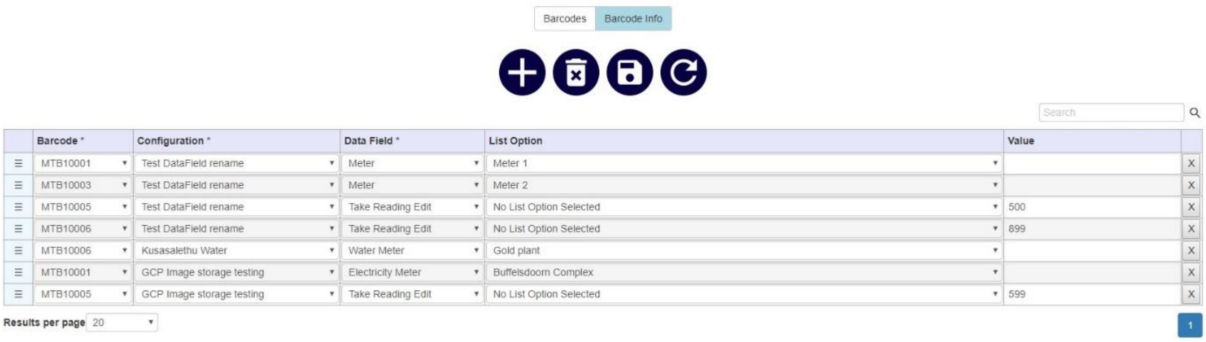
This section details the steps involved when adding **Barcode** scanning functionality to a **Configuration**. **Barcodes** can be scanned, and different **Data Fields** (values) can be auto-filled on the Data Collection page (application).

1. **Barcodes** should be added using the grid within the **Barcodes** tab within the **Barcode Management** menu.
    1. **Barcodes** can be added using a CSV file (Download a template using the Download Template grid button)
    2. The image below shows the *Barcodes Grid*

	Code *	
☰	MTB10001	X
☰	MTB10003	X
☰	MTB10005	X
☰	MTB10006	X

Figure A11: Barcodes Grid

2. **Barcodes** are linked to **Configurations** using the **Barcode Info Tab**
    1. Link an *existing* **Barcode** to a current **Configuration** for a specified **Data Field** using the grid.
    2. A **Barcode** should be connected to a **List Option** *or* a **Value**
    3. When a **Barcode** that is linked to a **List Option** is successfully scanned, a Data collection page (application) will automatically be opened with the **Data Field** associated with the linked **List Option** auto-filled with the related **List Option value.**
    4. When a **Barcode** that is linked to a **Value** is successfully scanned, a Data collection page (application) will automatically be opened with the linked **Data Field** auto-filled with the pre-configured **Value.**
    5. The image below shows the *Barcode Info Grid*



	Barcode *	Configuration *	Data Field *	List Option	Value	
☰	MTB10001	Test DataField rename	Meter	Meter 1		X
☰	MTB10003	Test DataField rename	Meter	Meter 2		X
☰	MTB10005	Test DataField rename	Take Reading Edit	No List Option Selected	500	X
☰	MTB10006	Test DataField rename	Take Reading Edit	No List Option Selected	899	X
☰	MTB10006	Kusasalethu Water	Water Meter	Gold plant		X
☰	MTB10001	GCP Image storage testing	Electricity Meter	Buffelsdoorn Complex		X
☰	MTB10005	GCP Image storage testing	Take Reading Edit	No List Option Selected	599	X

Figure A12: Barcode Info Grid

1.1.6 User Management

This section details the steps in providing users access to the Mobile Toolbox application. Mobile Toolbox users require an ATB account to log in to the application.

1. Use the **Users Database Manager** menu to add **Users** for the required group.
    1. Ensure that users have the **ATB - Access Mobile Toolbox Right**
2. **Users** are linked to **Configurations** and **Templates** using the **User Management** menu.
    1. When linking users to **Templates**, ensure that they are also linked to the matching **Configurations** (*parents* and *children*)

1.2 Generated report from consolidated data

Opening Metre Reading		Purchase Receipts			Metre Reading before Disposal	Disposal					Metre Reading After Disposal	Activity
Date	Opening Balance Reading (Litres)	Invoice Number	Date	Litres Received	Opening Balance plus Receipts	Date	Litres Disposed (if losses occurred, also note here)	DISPOSED TO Each vehicle/unit to be indicated separately		Purpose of Disposal (State whether or Eligible or non-eligible use)	Opening balance plus Receipts minus Disposals	Eligible/Non - Eligible
2020-11-01	65 870	n/a	n/a	0	65 870	2020-11-01	251	Grader	CAM4016	Scrape and level the road surfaces	65 619	E
2020-11-01	65 619	n/a	n/a	0	65 619	2020-11-01	164	Dump Truck	CAM1327	Cargo transport	65 455	E
2020-11-01	65 455	n/a	n/a	0	65 455	2020-11-01	38	Excavator	CAM2894	Load ore to dump trucks	65 417	E
2020-11-01	65 417	n/a	n/a	0	65 417	2020-11-01	200	Dump Truck	CAM1326	Cargo transport	65 217	E
2020-11-01	65 217	n/a	n/a	0	65 217	2020-11-01	51	LDVs (Bakkies)	CAM9274	Transport equipment and people	65 166	NE
2020-11-01	65 166	n/a	n/a	0	65 166	2020-11-01	70	LDVs (Bakkies)	CAM9322	Transport equipment and people	65 096	NE
2020-11-01	65 096	n/a	n/a	0	65 096	2020-11-01	69	LDVs (Bakkies)	CAM9305	Transport equipment and people	65 027	NE
2020-11-01	65 027	n/a	n/a	0	65 027	2020-11-01	589	Dump Truck	CAM1234	Cargo transport	64 438	E
2020-11-01	64 438	n/a	n/a	0	64 438	2020-11-01	387	Dump Truck	CAM1238	Cargo transport	64 051	E
2020-11-01	64 051	n/a	n/a	0	64 051	2020-11-01	435	Dump Truck	CAM1202	Cargo transport	63 616	E
2020-11-01	63 616	n/a	n/a	0	63 616	2020-11-01	502	Dump Truck	CAM1204	Cargo transport	63 114	E

Figure B1: Consolidated data report Extract 1

Opening Metre Reading		Purchase Receipts			Metre Reading before Disposal	Disposal					Metre Reading After Disposal	Activity
Date	Opening Balance Reading (Litres)	Invoice Number	Date	Litres Received	Opening Balance plus Receipts	Date	Litres Disposed (if losses occurred, also note here)	DISPOSED TO Each vehicle/unit to be indicated separately		Purpose of Disposal (State whether or Eligible or non-eligible use)	Opening balance plus Receipts minus Disposals	Eligible/Non - Eligible
2020-11-01	63 114	n/a	n/a	0	63 114	2020-11-01	183	Dump Truck	CAM1321	Cargo transport	62 931	E
2020-11-01	62 931	n/a	n/a	0	62 931	2020-11-01	331	Dump Truck	V39	Cargo transport	62 600	E
2020-11-01	62 600	n/a	n/a	0	62 600	2020-11-01	72	LDVs (Bakkies)	CAM9236	Transport equipment and people	62 528	NE
2020-11-01	62 528	n/a	n/a	0	62 528	2020-11-01	107	HIP-300	CAM5015	Transport equipment and parts	62 421	E
2020-11-01	62 421	n/a	n/a	0	62 421	2020-11-01	289	Dump Truck	V37	Cargo transport	62 132	E
2020-11-02	62 132	n/a	n/a	0	62 132	2020-11-02	127	Dump Truck	CAM1326	Cargo transport	62 005	E
2020-11-02	62 005	n/a	n/a	0	62 005	2020-11-02	160	Dump Truck	CAM1327	Cargo transport	61 845	E
2020-11-02	61 845	n/a	n/a	0	61 845	2020-11-02	263	Dump Truck	CAM1222	Cargo transport	61 582	E
2020-11-02	61 582	n/a	n/a	0	61 582	2020-11-02	108	LDVs (Bakkies)	CAM5028	Transport equipment and people	61 474	NE
2020-11-02	61 474	n/a	n/a	0	61 474	2020-11-02	317	Dump Truck	V34	Cargo transport	61 157	E
2020-11-02	61 157	n/a	n/a	0	61 157	2020-11-02	321	Dump Truck	V38	Cargo transport	60 836	E
2020-11-02	60 836	n/a	n/a	0	60 836	2020-11-02	312	Dump Truck	V40	Cargo transport	60 524	E

Figure B2: Consolidated data report Extract 2

Opening Metre Reading		Purchase Receipts			Metre Reading before Disposal	Disposal					Metre Reading After Disposal	Activity
Date	Opening Balance Reading (Litres)	Invoice Number	Date	Litres Received	Opening Balance plus Receipts	Date	Litres Disposed (if losses occurred, also note here)	DISPOSED TO Each vehicle/unit to be indicated separately		Purpose of Disposal (State whether or Eligible or non-eligible use)	Opening balance plus Receipts minus Disposals	Eligible/Non - Eligible
2020-11-02	60 524	n/a	n/a	0	60 524	2020-11-02	503	Dump Truck	CAM1198	Cargo transport	60 021	E
2020-11-02	60 021	n/a	n/a	0	60 021	2020-11-02	362	Dump Truck	CAM1321	Cargo transport	59 659	E
2020-11-03	59 659	n/a	n/a	0	59 659	2020-11-03	371	Dump Truck	ADT008	Cargo transport	59 288	E
2020-11-03	59 288	n/a	n/a	0	59 288	2020-11-03	532	Dump Truck	CAM1202	Cargo transport	58 756	E
2020-11-03	58 756	n/a	n/a	0	58 756	2020-11-03	555	Dump Truck	CAM1206	Cargo transport	58 201	E
2020-11-03	58 201	n/a	n/a	0	58 201	2020-11-03	199	Dump Truck	CAM1327	Cargo transport	58 002	E
2020-11-03	58 002	n/a	n/a	0	58 002	2020-11-03	212	Grader	CAM4027	Scrape and level the road surfaces	57 790	E
2020-11-03	57 790	n/a	n/a	0	57 790	2020-11-03	8 942	T - Diesel Bowser (transfer tank)	CAM1210	Deliver diesel to mining equipment	48 848	E
2020-11-03	48 848	n/a	n/a	0	48 848	2020-11-03	280	Dump Truck	CAM1221	Cargo transport	48 568	E
2020-11-03	48 568	n/a	n/a	0	48 568	2020-11-03	4 670	T - Diesel Bowser (transfer tank)	CAM5022	Deliver diesel to mining equipment	43 898	E
2020-11-03	43 898	n/a	n/a	0	43 898	2020-11-03	393	Dump Truck	V35	Cargo transport	43 505	E

Figure B3: Consolidated data report Extract 3

Opening Metre Reading		Purchase Receipts			Metre Reading before Disposal	Disposal					Metre Reading After Disposal	Activity
Date	Opening Balance Reading (Litres)	Invoice Number	Date	Litres Received	Opening Balance plus Receipts	Date	Litres Disposed (if losses occurred, also note here)	DISPOSED TO Each vehicle/unit to be indicated separately		Purpose of Disposal (State whether or Eligible or non-eligible use)	Opening balance plus Receipts minus Disposals	Eligible/Non - Eligible
2020-11-03	43 505	n/a	n/a	0	43 505	2020-11-03	299	Grader	CAM4016	Scrape and level the road surfaces	43 206	E
2020-11-03	43 206	n/a	n/a	0	43 206	2020-11-03	70	LDVs (Bakkies)	CAM9256	Transport equipment and people	43 136	NE
2020-11-03	43 136	n/a	n/a	0	43 136	2020-11-03	55	LDVs (Bakkies)	CAM9190	Transport equipment and people	43 081	NE
2020-11-03	43 081	n/a	n/a	0	43 081	2020-11-03	63	LDVs (Bakkies)	CAM9284	Transport equipment and people	43 018	NE
2020-11-03	43 018	n/a	n/a	0	43 018	2020-11-03	52	LDVs (Bakkies)	CAM9189	Transport equipment and people	42 966	NE
2020-11-03	42 966	n/a	n/a	0	42 966	2020-11-03	70	LDVs (Bakkies)	CAM9179	Transport equipment and people	42 896	NE
2020-11-03	42 896	n/a	n/a	0	42 896	2020-11-03	5 706	T - Diesel Bowser (transfer tank)	CAM1210	Deliver diesel to mining equipment	37 190	E
2020-11-03	37 190	n/a	n/a	0	37 190	2020-11-03	488	Dump Truck	CAM1204	Cargo transport	36 702	E
2020-11-03	36 702	n/a	n/a	0	36 702	2020-11-03	226	Dump Truck	CAM1202	Cargo transport	36 476	E

Figure B4: Consolidated data report Extract 4

Opening Metre Reading		Purchase Receipts			Metre Reading before Disposal	Disposal					Metre Reading After Disposal	Activity
Date	Opening Balance Reading (Litres)	Invoice Number	Date	Litres Received	Opening Balance plus Receipts	Date	Litres Disposed (if losses occurred, also note here)	DISPOSED TO Each vehicle/unit to be indicated separately		Purpose of Disposal (State whether or Eligible or non-eligible use)	Opening balance plus Receipts minus Disposals	Eligible/Non - Eligible
2020-11-03	36 476	n/a	n/a	0	36 476	2020-11-03	422	Dump Truck	CAM1198	Cargo transport	36 054	E
2020-11-03	36 054	n/a	n/a	0	36 054	2020-11-03	274	Dump Truck	CAM1205	Cargo transport	35 780	E
2020-11-03	35 780	n/a	n/a	0	35 780	2020-11-03	364	Dump Truck	CAM1321	Cargo transport	35 416	E
2020-11-03	35 416	n/a	n/a	0	35 416	2020-11-03	213	Dump Truck	ADT009	Cargo transport	35 203	E
2020-11-03	35 203	n/a	n/a	0	35 203	2020-11-03	183	Dump Truck	CAM1313	Cargo transport	35 020	E
2020-11-03	35 020	n/a	n/a	0	35 020	2020-11-03	301	Dump Truck	CAM1327	Cargo transport	34 719	E
2020-11-03	34 719	n/a	n/a	0	34 719	2020-11-03	324	Dump Truck	CAM1289	Cargo transport	34 395	E
2020-11-03	34 395	n/a	n/a	0	34 395	2020-11-03	8 171	T - Diesel Bowser (transfer tank)	CAM1210	Deliver diesel to mining equipment	26 224	E
2020-11-03	26 224	n/a	n/a	0	26 224	2020-11-03	2 039	T - Diesel Bowser (transfer tank)	CAM5022	Deliver diesel to mining equipment	24 185	E
2020-11-03	24 185	n/a	n/a	0	24 185	2020-11-03	256	Dump Truck	CAM1238	Cargo transport	23 929	E
2020-11-03	23 929	n/a	n/a	0	23 929	2020-11-03	69	LDVs (Bakkies)	CAM9235	Transport equipment and people	23 860	NE

Figure B5: Consolidated data report Extract 5

Opening Metre Reading		Purchase Receipts			Metre Reading before Disposal	Disposal					Metre Reading After Disposal	Activity
Date	Opening Balance Reading (Litres)	Invoice Number	Date	Litres Received	Opening Balance plus Receipts	Date	Litres Disposed (if losses occurred, also note here)	DISPOSED TO Each vehicle/unit to be indicated separately		Purpose of Disposal (State whether or Eligible or non-eligible use)	Opening balance plus Receipts minus Disposals	Eligible/Non - Eligible
2020-11-03	23 860	n/a	n/a	0	23 860	2020-11-03	25	LDVs (Bakkies)	CAM9305	Transport equipment and people	23 835	NE
2020-11-03	23 835	n/a	n/a	0	23 835	2020-11-03	63	LDVs (Bakkies)	CAM0874	Transport equipment and people	23 772	NE
2020-11-04	23 772	INV009036 0574	2020-11-04	31 997	55 769	2020-11-04	346	Dump Truck	CAM1222	Cargo transport	55 423	E
2020-11-04	55 423	n/a	n/a	0	55 423	2020-11-04	402	Dump Truck	ADT008	Cargo transport	55 021	E
2020-11-04	55 021	n/a	n/a	0	55 021	2020-11-04	545	Dump Truck	CAM1203	Cargo transport	54 476	E
2020-11-04	54 476	n/a	n/a	0	54 476	2020-11-04	5 182	T - Diesel Bowser (transfer tank)	CAM1210	Deliver diesel to mining equipment	49 294	E
2020-11-04	49 294	n/a	n/a	0	49 294	2020-11-04	160	Dump Truck	CAM1289	Cargo transport	49 134	E
2020-11-04	49 134	n/a	n/a	0	49 134	2020-11-04	443	Dump Truck	CAM1202	Cargo transport	48 691	E
2020-11-04	48 691	n/a	n/a	0	48 691	2020-11-04	306	Dump Truck	ADT009	Cargo transport	48 385	E
2020-11-04	48 385	n/a	n/a	0	48 385	2020-11-04	294	Dump Truck	CAM1321	Cargo transport	48 091	E
2020-11-04	48 091	n/a	n/a	0	48 091	2020-11-04	315	Dump Truck	CAM1204	Cargo transport	47 776	E
2020-11-04	47 776	n/a	n/a	0	47 776	2020-11-04	302	Dump Truck	CAM1238	Cargo transport	47 474	E

Figure B6: Consolidated data report Extract 6

Opening Metre Reading		Purchase Receipts			Metre Reading before Disposal	Disposal					Metre Reading After Disposal	Activity
Date	Opening Balance Reading (Litres)	Invoice Number	Date	Litres Received	Opening Balance plus Receipts	Date	Litres Disposed (if losses occurred, also note here)	DISPOSED TO Each vehicle/unit to be indicated separately		Purpose of Disposal (State whether or Eligible or non-eligible use)	Opening balance plus Receipts minus Disposals	Eligible/Non - Eligible
2020-11-04	47 474	n/a	n/a	0	47 474	2020-11-04	275	Dump Truck	CAM1313	Cargo transport	47 199	E
2020-11-04	47 199	n/a	n/a	0	47 199	2020-11-04	362	Dump Truck	CAM1287	Cargo transport	46 837	E
2020-11-04	46 837	n/a	n/a	0	46 837	2020-11-04	140	Diesel Bowser (engine tank)	CAM5022	Deliver diesel to mining equipment	46 697	E
2020-11-04	46 697	n/a	n/a	0	46 697	2020-11-04	353	Dump Truck	CAM1198	Cargo transport	46 344	E
2020-11-04	46 344	n/a	n/a	0	46 344	2020-11-04	282	Grader	CAM4027	Scrape and level the road surfaces	46 062	E
2020-11-04	46 062	n/a	n/a	0	46 062	2020-11-04	36	Excavator	CAM2894	Load ore to dump trucks	46 026	E
2020-11-04	46 026	n/a	n/a	0	46 026	2020-11-04	277	Grader	CAM4016	Scrape and level the road surfaces	45 749	E
2020-11-04	45 749	n/a	n/a	0	45 749	2020-11-04	74	Dump Truck	CAM1290	Cargo transport	45 675	E
2020-11-04	45 675	n/a	n/a	0	45 675	2020-11-04	8 026	T - Diesel Bowser (transfer tank)	CAM1210	Deliver diesel to mining equipment	37 649	E

Figure B7: Consolidated data report Extract 7

Opening Metre Reading		Purchase Receipts			Metre Reading before Disposal	Disposal					Metre Reading After Disposal	Activity
Date	Opening Balance Reading (Litres)	Invoice Number	Date	Litres Received	Opening Balance plus Receipts	Date	Litres Disposed (if losses occurred, also note here)	DISPOSED TO Each vehicle/unit to be indicated separately		Purpose of Disposal (State whether or Eligible or non-eligible use)	Opening balance plus Receipts minus Disposals	Eligible/Non - Eligible
2020-11-04	37 649	n/a	n/a	0	37 649	2020-11-04	2 531	T - Diesel Bowser (transfer tank)	CAM5022	Deliver diesel to mining equipment	35 118	E
2020-11-04	35 118	n/a	n/a	0	35 118	2020-11-04	235	Dump Truck	CAM1326	Cargo transport	34 883	E
2020-11-04	34 883	n/a	n/a	0	34 883	2020-11-04	204	Dump Truck	CAM1327	Cargo transport	34 679	E
2020-11-04	34 679	n/a	n/a	0	34 679	2020-11-04	58	Dump Truck	CAM9305	Cargo transport	34 621	E
2020-11-04	34 621	n/a	n/a	0	34 621	2020-11-04	18	Dump Truck	CAM9305	Cargo transport	34 603	E
2020-11-05	34 603	n/a	n/a	0	34 603	2020-11-05	218	Dump Truck	CAM1327	Cargo transport	34 385	E
2020-11-05	34 385	n/a	n/a	0	34 385	2020-11-05	421	Dump Truck	CAM1205	Cargo transport	33 964	E
2020-11-05	33 964	n/a	n/a	0	33 964	2020-11-05	74	LDVs (Bakkies)	CAM9094	Transport equipment and people	33 890	NE
2020-11-05	33 890	n/a	n/a	0	33 890	2020-11-05	50	LDVs (Bakkies)	CAM9317	Transport equipment and people	33 840	NE
2020-11-05	33 840	n/a	n/a	0	33 840	2020-11-05	5 129	T - Diesel Bowser (transfer tank)	CAM1210	Deliver diesel to mining equipment	28 711	E

Figure B8: Consolidated data report Extract 8

Opening Metre Reading		Purchase Receipts			Metre Reading before Disposal	Disposal						Metre Reading After Disposal	Activity
Date	Opening Balance Reading (Litres)	Invoice Number	Date	Litres Received	Opening Balance plus Receipts	Date	Litres Disposed (if losses occurred, also note here)	DISPOSED TO Each vehicle/unit to be indicated separately		Purpose of Disposal (State whether or Eligible or non-eligible use)	Opening balance plus Receipts minus Disposals	Eligible/Non - Eligible
2020-11-05	28 711	n/a	n/a	0	28 711	2020-11-05	69	HIP-300	CAM5015	Transport equipment and parts	28 642	E
2020-11-05	28 642	n/a	n/a	0	28 642	2020-11-05	402	Dump Truck	CAM1206	Cargo transport	28 240	E
2020-11-05	28 240	n/a	n/a	0	28 240	2020-11-05	56	LDVs (Bakkies)	CAM9274	Transport equipment and people	28 184	NE
2020-11-05	28 184	n/a	n/a	0	28 184	2020-11-05	184	Dump Truck	CAM1289	Cargo transport	28 000	E
2020-11-05	28 000	n/a	n/a	0	28 000	2020-11-05	58	LDVs (Bakkies)	CAM9147	Transport equipment and people	27 942	NE
2020-11-05	27 942	n/a	n/a	0	27 942	2020-11-05	307	Diesel Bowser (engine tank)	CAM1210	Deliver diesel to mining equipment	27 635	E
2020-11-05	27 635	n/a	n/a	0	27 635	2020-11-05	217	Dump Truck	CAM1326	Cargo transport	27 418	E
2020-11-05	27 418	n/a	n/a	0	27 418	2020-11-05	205	Dump Truck	CAM1290	Cargo transport	27 213	E
2020-11-05	27 213	n/a	n/a	0	27 213	2020-11-05	170	Dump Truck	CAM1327	Cargo transport	27 043	E
2020-11-05	27 043	n/a	n/a	0	27 043	2020-11-05	166	Dump Truck	CAM1287	Cargo transport	26 877	E
2020-11-05	26 877	n/a	n/a	0	26 877	2020-11-05	166	Dump Truck	CAM1313	Cargo transport	26 711	E

Figure B9: Consolidated data report Extract 9

Opening Metre Reading		Purchase Receipts			Metre Reading before Disposal	Disposal						Metre Reading After Disposal	Activity
Date	Opening Balance Reading (Litres)	Invoice Number	Date	Litres Received	Opening Balance plus Receipts	Date	Litres Disposed (if losses occurred, also note here)	DISPOSED TO Each vehicle/unit to be indicated separately		Purpose of Disposal (State whether or Eligible or non-eligible use)	Opening balance plus Receipts minus Disposals	Eligible/Non - Eligible
2020-11-30	54 366	n/a	n/a	0	54 366	2020-11-30	73	HIP-300	CAM5015	Transport equipment and parts	54 293	E
2020-11-30	54 293	n/a	n/a	0	54 293	2020-11-30	8 939	T - Diesel Bowser (transfer tank)	CAM1210	Deliver diesel to mining equipment	45 354	E
2020-11-30	45 354	n/a	n/a	0	45 354	2020-11-30	36	LDVs (Bakkies)	CAM9310	Transport equipment and people	45 318	NE
2020-11-30	45 318	n/a	n/a	0	45 318	2020-11-30	69	LDVs (Bakkies)	CAM9163	Transport equipment and people	45 249	NE
2020-11-30	45 249	n/a	n/a	0	45 249	2020-11-30	230	Watercart	CAM1122	Dust control	45 019	E
2020-11-30	45 019	n/a	n/a	0	45 019	2020-11-30	461	Dump Truck	CAM1198	Cargo transport	44 558	E
2020-11-30	44 558	n/a	n/a	0	44 558	2020-11-30	298	Grader	CAM4027	Scrape and level the road surfaces	44 260	E
Opening Balance	65 870	Plus	Total Received	542 256	Minus	Total Disposed	563 866	Equals	44 260	closing balance	40 706	
										Estimated Losses (Litres)	3 554	

Figure B10: Consolidated data report Extract 10