FEASIBILITY STUDY OF IMPLEMENTING A NATIONAL HEALTH DATA REPOSITORY IN RWANDA TO IMPROVE PATIENT DATA ACCESSIBILITY AND HEALTHCARE SERVICE DELIVERY.

CASE STUDY OF REFERRAL AND DISTRICT HOSPITALS

By

NSENGIYUMVA Emmanuel

Registration number: 215027572

A dissertation submitted to University of Rwanda in a partial fulfillment of the requirements for the degree of

MASTER OF HEALTH INFORMATICS

In the College of Medicine and Health Sciences

Regional Centre of Excellence in Biomedical Engineering and e-Health

Supervisor: Prof. Marc Nyssen

Co-Supervisor: Associate Prof. NIYOYITA Jean Paul

September, 2023

DECLARATION

I, NSENGIYUMVA Emmanuel, declare that this research proposal is titled **"Feasibility study of Implementing a national health data repository in Rwanda to improve patient data accessibility and healthcare service delivery. Case study of referral and district hospitals"** submitted in fulfilment of University of Rwanda academic rules and regulations for the master's degree in health informatics; is our original work and it has never been presented elsewhere (in any universities or institution of higher learning) for any purpose. And I do declare that all cited information is shown in the provided list of references as indicators of the sources. The publication is granted to the University of Rwanda/College of Medicine and Health Sciences for academic interest.

NSENGIYUMVA Emmanuel

Signature: 

Date: 12 September 2023

Supervisor: Prof. Marc Nyssen

Signature: 

Date: 12 September 2023

Co-Supervisor: Associate Prof. NIYOYITA Jean Paul, PhD

Signature: 

Date: 13 September 2023

DEDICATION

This work is warmly dedicated to the Almighty God for the gift of life and love, our families especially our lovely parents, our sisters, brothers, and our friends for their extraordinary support and care. In addition, I am dedicating this work to my charming classmates with whom I have been together within my studies all along. Finally, I would like to dedicate this task to my lecturers, supervisors, and career guides I met during my studies for the knowledge, skills, and attitude that they have shared.

ACKNOWLEDGEMENT

I am very grateful and extend my lovely appreciation to the almighty God for his mercy, blessing, and by whom this work has been made possible. I acknowledge with gratitude the contributions of the individuals, lecturers, friends, and families.

I would like to express my gratitude to my Supervisors Prof. Marc Nyssen and Ass. Prof. Niyoyita Jean Paul.

In addition, I would like to thank the staff of the university of Rwanda, the department of Health Informatics, the college of medicine and health sciences (CMHS) especially the Center of Excellent in Biomedical Engineering and e-Health (CEBE), and the government of Rwanda who facilitates in support of my entire study and giving me the different opportunities to get these skills. May God Bless you all.

ABSTRACT

Background: This comprehensive feasibility study aims at assessing the practicality and potential impact of establishing a centralized repository to store and manage healthcare data across the country. By facilitating improved patient data accessibility and healthcare service delivery, such an initiative holds the promise of revolutionizing the healthcare landscape in Rwanda.

Methods: The study design was cross-sectional conducted using a quantitative approach, which involves collecting and analysing numerical data to identify patterns. The study was conducted in two Kigali-based national referral hospitals namely King Faisal Hospital, and Ndera Neuropsychiatric Teaching Hospital. One Level two teaching hospital (Nyamata Level two Teaching Hospital) and two district hospitals (Gatonde District Hospital, and Nyarugenge District Hospital). Two sample groups were selected. The first sample group consisted of doctors who record patient data and use it in decision-making whereas the second sample group consisted of ICT staff and managing directors. The study sample included 100 healthcare providers and 20 ICT. Therefore, the total study sample was 120 participants.

Results: Support for a national health data repository is strong (66.7%) and slight (30%), with only 3.3% opposing it. Notably, "Makes patient information more accessible, and reusable" garnered the highest level of endorsement, with 94.17% of participants recognizing its potential. Afterward, "Enables health information exchange" garnered a significant level of support, with 86.67% of participants. Remarkably, “Data security and privacy concerns” garnered the highest level of endorsement, with a significant 93.33% of participants acknowledging its potential as a major obstacle. Lastly Among the surveyed individuals, a substantial majority, comprising 78.30%, favoured the utilization of a central data repository interconnected with all participating hospitals through the Extract, Transform, Load (ETL) Technology.

Conclusion: This repository could be a game-changer in healthcare, revolutionizing how patient information is managed and utilized to provide more efficient and effective medical care nationwide.

Keywords: Data repository, Data accessibility, Data interoperability, Patient information exchange

CONTENTS

Table of figures

List of Abbreviations

Abbreviation	Definition
UR	University of Rwanda
CMHS	College of Medicine and Health Sciences
CEBE	Center of Excellent in Biomedical Engineering and e-Health
ICT	Information and Communication Technologies
NHIRD	National Health Information Repository and Data Warehouse
NHI	National Health Insurance
HISP	Health Information Services Provider
HMIS	Health Management Information System
SISCOM	Community Health Information System
MOH	Ministry of Health
MINICT	Ministry of Information Communications Technology and Innovation
PIH	Partners in Health
EMR	Electronic Medical Record
EHR	Electronic Health Records
GPS	Global Positioning System
OCLC	Online Computer Library
DoS	Denial Of Service
HDR	Healthcare Data Repository
CDR	Clinical Data Repository
NHDR	National Health Data Repository
NHIRD	National Health Information Repository and Data system
UHC	Universal Health Coverage

CDW	Clinical Data Warehouse
EHDS	European Health Data Space
NIH	National Institutes of Health
BTRIS	Biomedical Translational Research Information System
MHDR	Military Health System Data Repository
DHA	Defense Health Agency
EAV	Entity-Attribute-Value
CCD	Continuity of Care Documents
HIE	Health Information Exchange
NoSQL	Not Only SQL
SQL	Structured Query Language
M-R	MapReduce
YARN	Yet Another Resource Negotiator
LDW	Logical Data Warehouse
IRD	Institutional Review Board
ELT	Extract, Load and Transform,
ETL	Extract, Transform and Load

CHAPTER 1: INTRODUCTION

1.1. Introduction

This comprehensive feasibility study aims to investigate the practicality and potential impact of establishing a centralized repository to store and manage healthcare data across the country. By facilitating improved patient data accessibility and healthcare service delivery, such an initiative holds the promise of revolutionizing the healthcare landscape in Rwanda. This introduction provides an overview of the critical objectives and significance of the feasibility study, underscoring its importance in shaping the future of healthcare in the nation.

1.2. Background and Rationale

According to an International Data Corporation (IDC) report healthcare data amounted to 153 exabytes in 2013, and it was estimated to reach 2,314 exabytes by 2020. Hence, the fundamental difficulty faced by healthcare institutions lies in integrating data from diverse clinical systems, a crucial step toward enhancing patient care and performance metrics(1).

As reported in the 2019 Health Systems Trust report, the South African National Health Department has acknowledged that despite the various information sources accessible to managers, data remains challenging to access, often lacks integration, and consequently isn't utilized to the extent needed for informed decision-making. HISP-SA has effectively formulated, designed, created, and implemented the National Health Information Repository and Data Warehouse (NHIRD) system, a web-based repository for health information. This system aids in gathering information for the advancement of the National Health Insurance (NHI) system(2).

In recent years, the performance of the healthcare system in Rwanda has displayed noticeable enhancements, attributed to the enhanced quality of care and the decentralization of healthcare systems. Policies and initiatives have been crafted to address the healthcare requirements of the populace and to synchronize the national health system with the worldwide health objectives. Health data from health facilities and local communities are systematically transmitted by data managers and community health workers. Regular reporting occurs at specified intervals (quarterly, monthly, or weekly) or on an as-needed basis, utilizing the web-based Rwandan Health Management Information System (HMIS) (3).

The primary source of health data in Rwanda is the Health Management Information System (HMIS). Health centers and hospitals record data in various Electronic Medical Record systems (such as

OpenClinic, OpenMRS), which are structured into records within patient files. These records are subsequently compiled into reports on a monthly basis and transmitted to the central HMIS server by the fifth day of the reporting month (4). Additional regular web-based solutions for information management encompass the Integrated Disease Surveillance and Response system (responsible for monitoring various diseases like HIV and tuberculosis), data warehouse systems intertwined with performance-based financing, and the Community Health Information System (SISCOM). SISCOM gathers, stores, retrieves, and disseminates crucial program and patient information concerning care and treatment. The integration of data-driven decision-making and policy formulation has bolstered the efficiency of health program administration while enhancing the government's ability to oversee healthcare quality. The initiation of a centralized, web-based Health Management Information System (HMIS) commenced in 2012 with the aim of streamlining data collection processes and enhancing the timeliness of reporting (3).

In Rwanda, the Ministry of Health (MOH) joined forces with the Ministry of Information Communications Technology and Innovation (MINICT) to undertake a significant endeavor: the digitization of all healthcare services. This initiative aims to increase digital capacities for the intelligent acquisition, processing, and exchange of health data. The ultimate goal is to enhance the digital landscape of healthcare, enrich patient experiences, and elevate health outcomes within the country.(5) In 2005, Partners in Health (PIH) introduced Rwanda's first electronic medical record (EMR) system. The purpose was to enhance the quality of care for HIV and TB patients. This system contains comprehensive patient information, accessible to medical professionals either through printed reports or directly on computers within consultation rooms.(6)

Currently, different health systems are being implemented in Rwanda, including OpenMRS, OpenClinic, Babyl, Rapid SMS, HMIS, m-Health, TracPlus, and TRACnet, and other. This collection of systems has the objective of addressing the healthcare requirements of the population and enhancing the efficiency of health system administration to achieve greater productivity (7). All the above systems generate health data daily; the generated data play several important roles in healthcare service delivery but most of them work independently at the healthcare facilities level.

This study seeks to examine whether it is needed, practical and possible to establish the mentioned repository. It was delved into various aspects, including the technical requirements, financial considerations, legal and ethical implications, as well as the readiness of existing healthcare facilities to

embrace and integrate the new system. The findings of this study are anticipated to inform key stakeholders, policymakers, and healthcare administrators in making well-informed decisions to optimize patient care, enhance healthcare service delivery, and promote overall health outcomes for the Rwandan population. Ultimately, the successful implementation of a national health data repository in Rwanda could mark a significant step towards achieving a more efficient, patient-centric, and data-driven healthcare system.

1.3. Problem Statement

The healthcare sector has conventionally been rapid in embracing new technologies, yet it tends to be slow in addressing data-related matters, particularly concerning data sharing and integration. The overall situation can be related to a puzzle, with several components that need to fit together. Beyond the technical hurdles associated with integrating clinical data, there exists a challenge linked to the willingness and capacity for cooperation among stakeholders, encompassing healthcare providers and patients. Consequently, processes related to data collection, storage, integration, and analysis remain fragmented. The primary purpose of gathering healthcare data is to enhance the cost-effectiveness and quality of medical services while producing valuable insights for both patients, researchers, and decision makers. However, a fundamental question remains regarding the consolidation of integrated data originating from diverse formats and exhibiting varying degrees of quality and accessibility. This complexity arises due to the diverse array of systems employed within clinics as sources of such data. (1).

In Rwanda, a variety of health systems are used in healthcare facilities. Each of these systems works on the facility level and saves data on its saver. These systems are developed based on different Interoperability Standards. As result, it is tough for doctors, laboratories, pharmacies, and patients to share information, regardless of the specific tools they use. This poses a significant hurdle when it comes to exchanging health-related information between different healthcare facilities. However, effectively handling, accessing, and utilizing health data and information is crucial for understanding how healthcare services are used and assessing the quality of care delivered to the public. This, in turn, helps enhance the way we measure the effectiveness of healthcare and its outcomes. So, managing, accessing, and using health data are essential tasks that healthcare institutions and research organizations must focus on strategically (8)

In Rwanda, patients go through a lengthy process that involves different healthcare facilities. It all begins with local health workers and goes on through Health Posts, Health Centres, District Hospitals, Provincial Hospitals, Referral Hospitals, and Specialist Hospitals. However, the issue lies in the fact that there's a problem with getting hold of and sharing health information effectively. This means that patient healthcare records end up being recorded multiple times at each level. Because there isn't a single, unified patient record, healthcare providers face a big challenge. They must make important decisions about patient care with only limited access to their medical history. As a result, the lack of accessible and complete health information creates serious problems for patients. It can lead to healthcare services that are inefficient, ineffective, and even potentially harmful. This situation doesn't just affect the quality and safety of healthcare but also weakens the entire healthcare system's effectiveness in Rwanda.

What is crucial for maintaining the quality and effectiveness of healthcare is the ability to keep track of patients as they navigate through the various stages of the healthcare system, starting from basic healthcare and moving on to specialized care, hospital stays, long-term care, care at home, clinical treatments, and ultimately, end-of-life care. These records should also include details about the patient's personal characteristics, illnesses, medications, treatments, tests, and even medical images. This kind of ongoing monitoring provides a comprehensive overview of the healthcare services given and the resulting health outcomes. It is a way to identify medical mistakes, adverse reactions to medications, adherence to medical guidelines, effective therapies, optimal care plans, and the most successful responses to treatment. To truly understand these paths, it's necessary to link different sets of data at the level of individual patients, since current health information tends to be scattered across various healthcare facilities (9).

The concept of creating a data repository from electronic health records (EHRs) holds immense potential. It could lead to a fundamental leap in evaluating healthcare quality and performance, as these records could be consolidated into a national health data repository. This unified system would encompass patients' healthcare journeys and their outcomes, allowing for data extraction. However, progress in linking and extracting data from clinical record systems to establish a national health data repository is limited in many countries, including Rwanda (10).

1.4. Research objectives

1.4.1. Research Aim

This study aims to assess the feasibility and practicality of establishing a national health data repository in Rwanda to enhance patient data accessibility and elevate healthcare service delivery standards across the country.

1.4.2. Specific objectives

1. To assess the needs and level of support among participants for the implementation of a national healthcare data repository and the adoption of a Single Patient Record system.
2. To analyse the clinical benefits that can be derived from the utilization of a national health data repository and the implementation of a Single Patient Record system.
3. To identify the best approaches for the effective implementation of a national healthcare data repository.

1.5. Research questions

1. What are the specific needs and preferences of healthcare stakeholders (e.g., healthcare providers, Data manager and IT) regarding the implementation of a national healthcare data repository and the adoption of a Single Patient Record system?

2. What clinical benefits can be attributed to the utilization of a national healthcare data repository and the implementation of a Single Patient Record system, and how do these benefits impact patient care, healthcare outcomes, and overall healthcare efficiency?

3. What are the key strategies and best practices for the successful and efficient implementation of a national healthcare data repository, taking into consideration factors such as data security, data quality, interoperability, scalability, and stakeholder engagement?

1.6.Significance of the study

Establishing a nationwide health data repository in Rwanda carries immense significance in shaping the path ahead for the country's healthcare system. The true importance of this study is found in its potential to bring about meaningful change, greatly enhancing how patient data is accessed, reported, and ultimately, how healthcare services are provided. In addition, a centralized health data repository would provide healthcare professionals with a comprehensive and up-to-date view of patients' medical histories. This accessibility to complete patient data would lead to more accurate diagnoses, personalized

treatment plans, and improved patient outcomes. It would empower clinicians to make well-informed decisions, ensuring a higher standard of care across all levels of the healthcare system.

In an era where data plays a pivotal role in shaping healthcare practices, the exploration of a national health data repository holds immense promise for revolutionizing healthcare systems. This study delves into the significance of implementing a centralized repository for healthcare data in Rwanda. The potential impact spans multiple dimensions, from improved efficiency and data-driven decision-making to enhanced research capabilities and data security. Below are benefits of having a national health data repository and a single patient record system:

Improved Healthcare Efficiency: A central tenet of the proposed national health data repository lies in its potential to elevate healthcare efficiency. By unifying patient data from various healthcare facilities, redundant data entry and administrative burdens can be minimized. This streamlining of data management processes liberates valuable resources that can be redirected toward patient care, where they are most needed. As the burden of paperwork is eased, healthcare professionals can devote more time and attention to delivering quality medical care.

Data-Driven Decision Making: The implementation of a national health data repository can bring healthcare systems closer to realizing the dream of evidence-based decision-making. At the patient level, healthcare providers gain access to comprehensive medical histories, enabling them to make informed treatment decisions. On a broader scale, policymakers and administrators benefit from aggregated data that identifies health trends and patterns. This, in turn, enables them to allocate resources strategically and design targeted interventions to address prevailing health challenges with precision.

Interoperability and Information Exchange: A paramount advantage of a centralized repository is its role in facilitating interoperability and seamless information exchange between healthcare facilities. This interconnectedness fosters collaboration and coordination across the healthcare spectrum. It expedites patient referrals, ensures continuity of care, and bridges the gaps between disparate healthcare providers, leading to a more holistic approach to patient well-being.

Research and Innovation: In the realm of medical research, a national health data repository serves as a treasure trove of information. Researchers gain access to a rich and diverse dataset, primed for conducting large-scale studies, clinical trials, and epidemiological research. The resulting advancements in medical knowledge fuel innovation and drive healthcare practices forward, ultimately benefiting patients and the entire healthcare ecosystem.

Data Security and Privacy: Addressing concerns around data security and patient privacy is paramount to any data repository initiative. A rigorous feasibility study would lay the foundation for robust data protection mechanisms. Building trust among patients, healthcare professionals, and stakeholders ensures that sensitive health information is handled and utilized with the utmost care, safeguarding both individual privacy and overall data integrity.

Long-term Sustainable Healthcare: The potential of a national health data repository extends beyond the immediate benefits. By creating a cohesive and efficient healthcare data ecosystem, Rwanda can fortify the long-term sustainability of its healthcare system. This repository serves as a bedrock for continuous improvement, constant monitoring, and informed evaluation of healthcare practices. The result is a healthcare infrastructure that is agile, adaptive, and responsive to the evolving needs of the population.

In summation, this feasibility study embarks on a journey to reshape Rwanda's healthcare landscape. Through the lens of a national health data repository, it addresses data accessibility challenges, optimizes service delivery, and propels the nation toward a future of data-driven healthcare excellence. This endeavor has the potential to elevate patient care, empower healthcare professionals, and shape a healthier tomorrow for Rwanda.

1.7. Definition of Key terms

A Feasibility study: is a preliminary examination of a proposed project or endeavor to ascertain its merits and viability. A feasibility study is intended to give an unbiased evaluation of all aspects of a proposed project, including technical, economic, financial, legal, and environmental factors (11).

Data repository: A data repository is a centralized storage location for data that is used for reporting, analysis, and decision-making. It is a collection of data sets that are organized and managed in a way that makes it easy to find and use the data. Data repositories are designed to efficiently store and facilitate the retrieval, sharing, and analysis of data, making them valuable resources for various research, analysis, and operational purposes (12).

Data Accessibility: Data accessibility refers to the ease and availability with which data can be retrieved, retrieved, and used by authorized individuals or systems. It indicates the extent to which data can be conveniently accessed, shared, and utilized for various purposes, such as analysis, decision-making, research, or reporting. Data accessibility is crucial for efficient and effective data utilization, enabling timely and informed actions based on the information stored in databases or repositories (13).

Single Patient Record: A Single Patient Record refers to a comprehensive and unified electronic or digital compilation of an individual's medical and healthcare information. This record consolidates various types of data, including medical history, diagnoses, treatments, medications, test results, imaging, and other relevant healthcare-related details. The purpose of a Single Patient Record is to provide healthcare providers with a holistic view of a patient's health status, enabling better coordination of care, more accurate diagnoses, and improved decision-making in the delivery of medical services (14).

CHAPTER 2. LITERATURE REVIEW

The implementation of a national health data repository has emerged as a promising solution to enhance patient data accessibility and optimize healthcare service delivery in Rwanda. As the country continues its efforts to improve healthcare systems and provide better patient care, the need for a centralized repository to efficiently manage and exchange healthcare data has become increasingly apparent (2). This literature review delves into the feasibility study of establishing a national health data repository in Rwanda, focusing on the specific case study of referral and district hospitals. By examining existing research, scholarly articles, and expert insights, this review aims to explore the potential benefits, challenges, and implications of such an initiative in revolutionizing Rwanda's healthcare landscape.

By synthesizing existing literature and insights on the feasibility study of implementing a national health data repository, this review aimed to provide a comprehensive understanding of the implications and significance of such an initiative for Rwanda's healthcare system. Ultimately, the findings of this literature review can inform policymakers, healthcare administrators, and stakeholders in making informed decisions that optimize patient care and healthcare service delivery, leading to improved health outcomes and more resilient healthcare infrastructure in Rwanda.

2.1. Empirical review

The potential enhancements in healthcare and health research achieved through data-intensive applications depend on the increasing volume of health data. Within the extensive endeavor of integrating data on a large scale, clinical data warehouses (CDWs) play a pivotal role. They not only handle data governance but also oversee data access and reuse. With the growing complexity of data management, it becomes essential to enhance transparency and standardize the criteria and processes. This ensures objective oversight and control. Consequently, the development of policies grounded in practicality and supported by evidence is of paramount importance. This study aimed to evaluate the range of criteria and procedures governing data access and utilization in clinical data warehouses on an international scale.(15)

A data repository denotes a comprehensive storage system, often taking the form of a large database infrastructure or a curated assemblage of databases. Its primary function involves collecting, managing, and storing data alongside corresponding metadata. This arrangement facilitates diverse operations such as data access, analysis, reporting, and data sharing. Data is amassed and stored in varying manners,

including aggregated data which stems from multiple sources or segments of a business. These data can be organized in structured or unstructured formats and subsequently enriched with distinct metadata (16). An inherent feature of data repositories is their capacity to search and extract information across extensive datasets. The extracted data can then be subjected to analysis, yielding potentially substantial efficiencies across various domains (17).

In 2022, the European Commission initiated the European Health Data Space (EHDS), a central pillar in the establishment of a robust European Health Union. This initiative provides a consistent, reliable, and efficient framework for utilizing health data in research, innovation, policy development, and regulatory activities. Importantly, it ensures full compliance with the European Union's rigorous data protection standards. EHDS also empowers individuals to manage and utilize their health data, both within their home country and across EU Member States, fostering a unified market for digital health services and products (18).

In the United States, a similar effort led to the creation of the United States National Institutes of Health (NIH), which encompasses 27 institutes and centers (ICs) dedicated to advancing biomedical research for the betterment of public health. In 2009, it evolved into the NIH Biomedical Translational Research Information System (BTRIS), boasting a substantial database and employing innovative data models and terminology management techniques. BTRIS plays a pivotal role in merging data from diverse sources, enabling active clinical trials, and facilitating the reuse of valuable healthcare information (19).

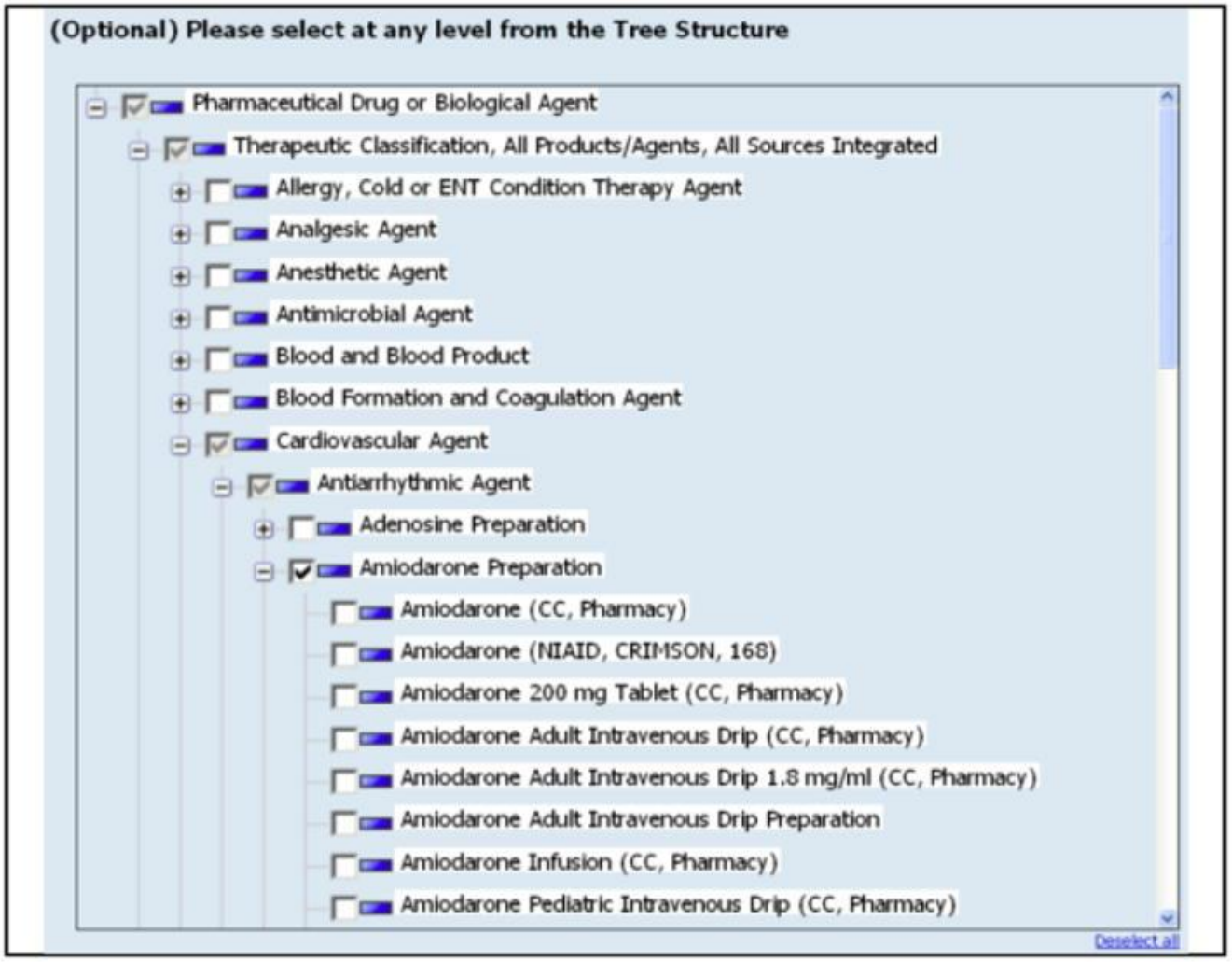

Figure 1. Example of tree-based terminology search in BTRIS.

BTRIS was conceived as a comprehensive repository for all clinical research data gathered from subjects at the NIH. This journey commenced with initial assumptions regarding design prerequisites and preferred features, and swiftly progressed through the stages of design, construction, and implementation. Our innovative approach, which combines elements from both column-oriented and Entity-Attribute-Value (EAV) data models, has empowered us to effectively handle a wide array of data from various sources, enabling seamless data aggregation across multiple sources. This fusion of column-oriented and EAV data models has provided healthcare professionals with the capability to manage diverse data originating from disparate sources, facilitating data aggregation across numerous data repositories (19).

Additionally, within the military domain, there exists a national health data repository known as the Military Health System Data Repository (MHDR). The MHDR serves as the central data repository responsible for capturing, preserving, and disseminating healthcare data on a global scale for the Defense Health Agency (DHA), including Tricare. It performs data reception and validation from a vast network of over 260 healthcare facilities worldwide within the Department of Defense (DoD) and from external, non-DoD data sources. This comprehensive coverage encompasses approximately 9.2 million beneficiaries. The MHDR employs data quality checks to enhance the usefulness of DHA's corporate data and offers both online and near-line data storage capabilities, facilitating the secure transfer of healthcare data (19).

In the year 2022, the Ministry of Health in Philippine put forth comparable proposals to establish a National Health Data Repository (NHDR). This repository is designed to enable the accessibility of healthcare services and health-related information, while ensuring the secure sharing and exchange of data. The overarching goal was to improve the safety, quality, fairness, and responsiveness of the healthcare system for the entire Filipino population. This transformation represents a significant step towards revolutionizing the use of information for the planning, management, delivery, and surveillance of healthcare services (20). The following is the proposed National Health Data Repository Framework Philippine:

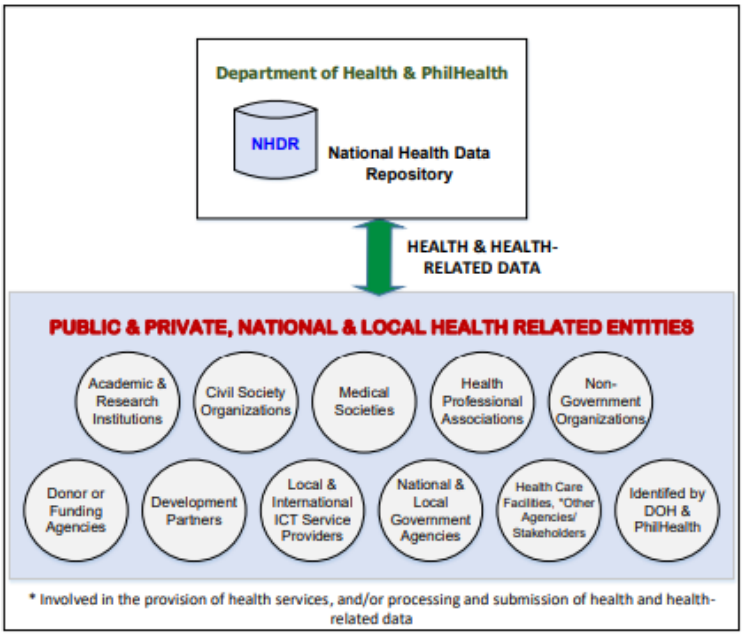

Figure 2. Proposed National Health Data Repository Framework Philippine

The overarching aim of the NHDR was to establish and execute a comprehensive approach encompassing data submission, processing, management, sharing, access, and governance for health and health-related data, all geared toward evidence-driven sectoral policymaking and Universal Health Coverage (UHC) planning. This initiative seeks to address persistent challenges related to the quality and timeliness of health data reporting, ensuring that high-quality health data and reports are readily accessible to stakeholders while adhering to legal, ethical, secure, consistent, and efficient data processing practices across all healthcare levels. It also aims to enhance health system performance while safeguarding data protection, enabling systematic collection, analysis, and timely dissemination of precise epidemiologic surveillance data for public health program planning, implementation, and evaluation. Additionally, it facilitates real-time data collection and analysis concerning the utilization of the Special Health Fund and the compilation and publication of price data for various health services and goods (20).

Around the world, numerous countries are embracing similar initiatives aimed at establishing national health data repositories. In the case of South Africa, this endeavor was initiated in 2019 and spans a five-year period, concluding in 2024. A publication from the country's Ministry of Health highlighted the transformative potential of digital health technologies in enhancing health systems and reshaping the delivery of healthcare services, as well as interactions between individuals and these services. The National Digital Health Strategy for South Africa, in effect from 2019 to 2024, sought to strengthen governance structures for digital health, create robust integrated platforms for information systems development, and collaborate with other government departments to establish essential broadband network infrastructure (21).

At the heart of this strategy was a vision of 'Enhanced Health for all South Africans through person-centered Digital Health.' The strategy brought forth numerous advantages, benefiting patients in their quest for healthcare access, healthcare professionals striving to provide improved services, and health system managers in fulfilling their duties. This initiative empowered all citizens to navigate their personal health journeys more effectively using digital technologies. The project had been long overdue, with legislative processes, including the passing of various laws related to the national health repository, playing a significant role in its development and implementation (21).

Table 1.A selection of legislation that has an impact on digital health.

Act of Parliament	Purpose
Electronic Communications Security (Pty) Ltd Act, 2002 (Act 68 of 2002)	Privacy and confidentiality of communications
Electronic Communications and Transactions Act, 2002 (Act 25 of 2002)	Privacy and confidentiality of communications
Electronic Communications Act, 2005 (Act 36 of 2005)	Basis for universal connectivity
Protection of Personal Information Act, 2013 (Act 4 of 2013)	Privacy and confidentiality of health records and how they are handled
The National Archives of South Africa Act, 1996 (Act 43 of 1996)	Defines a set of security protocols and retention requirements for any health record, including electronic records

The White Paper for NHI Policy in 2017 called for the creation of an integrated and enhanced national health information repository and data system (NHIRD) to effectively manage the NHI initiative. The NHIRD was specifically designed to collect a defined set of data from the National Electronic Health Record (EHR). Its primary purpose was to provide essential data and analytical insights to support the National Health Insurance Fund (NHIF) and played a crucial role in the implementation of the NHI plan, ensuring smooth access to healthcare services for the population.

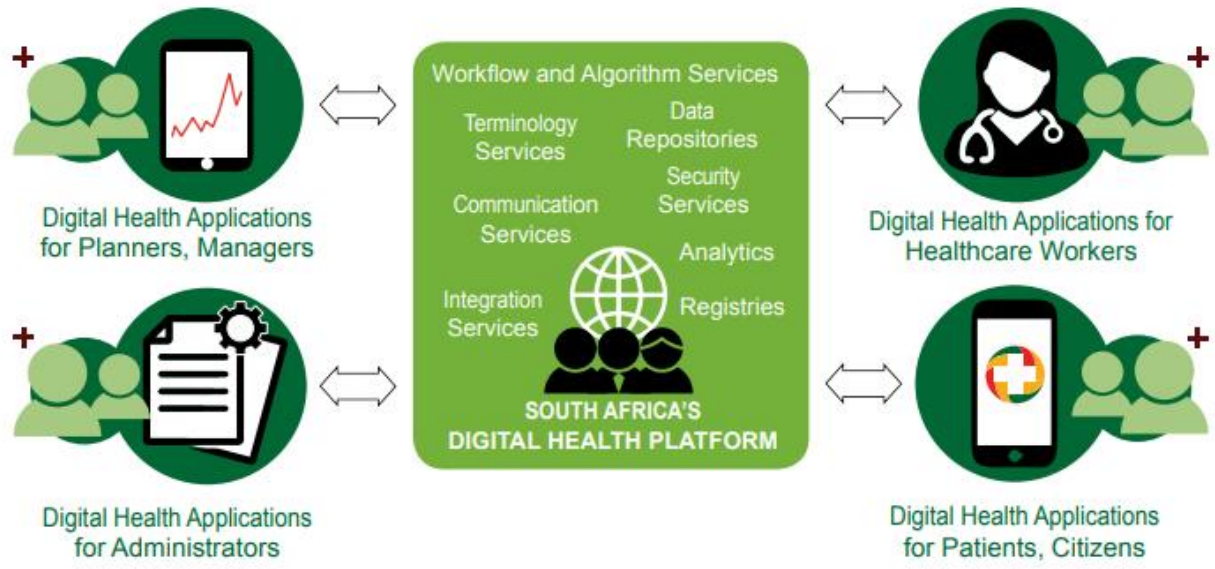

Figure 3. Conceptual overview of South Africa's Digital Health Platform.

2.2. Types of data repository

Data repositories are generally categorized into four types of data repositories.

Data Warehouse: The Data Warehouse represents the most expansive category of data repositories, encompassing data derived from diverse business sectors or sources such as transactional systems, relational databases, application log files, devices, APIs, and other comparable sources. This data is typically gathered at regular intervals. Within this repository, the stored data predominantly serves the purpose of analysis and reporting, facilitating informed decision-making for users and teams engaged in various business endeavors or projects. A distinctive advantage of data warehouses lies in their capacity to analyse extensive volumes of heterogeneous data, extracting significant insights and retaining historical records.

From a broader viewpoint, a data warehouse offers a unified perspective of either a physical or logical data repository consolidated from multiple systems. The primary objective of a data warehouse revolves around establishing cohesion among data extracted from existing systems. For instance, patient clinical data might be housed in one Electronic Medical Record (EMR) system, while their medication records reside in another. (22)

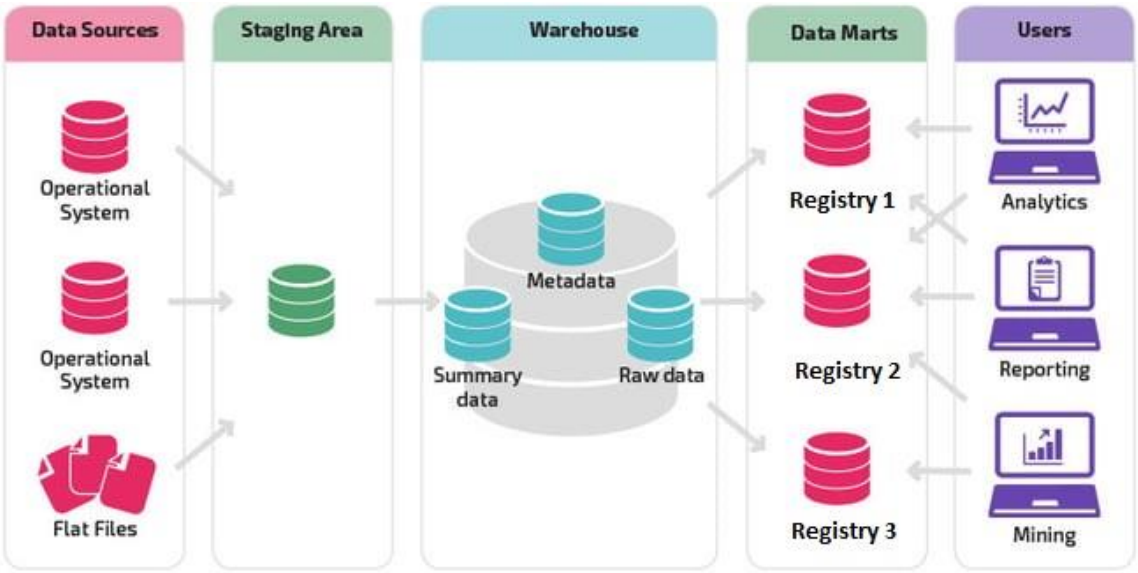

Figure 4. Data Warehouse Architecture

Data Lake: The Data Lake model entails a repository where data can exist in a multitude of formats, ranging from structured to semi-structured and unstructured arrangements. The fundamental motivation behind the inception of the data lake concept is in response to the limitations encountered with conventional data warehouses. This approach offers an avenue for refined data governance practices and the attainment of comprehensive data control. Data within this repository retains its raw, unprocessed

nature, making it amenable to a wide array of applications, including reporting, visualization, advanced analytics, and the deployment of machine learning methodologies. (23)

Data Mart: This can be perceived as a segment within the data warehouse that centers around a particular subject, department, or distinct domain. With data curated specifically for a given scope, users can efficiently retrieve insights without the need to extensively search through the entire expanse of the data warehouse. Ultimately, this expedites access and simplifies the user experience.(22)

Data Cube: This storage system holds the most intricate data. Think of it as an advanced version of tables with multiple dimensions. It's typically utilized to represent data that is too intricate to be captured solely through rows and columns in traditional tables. Essentially, a data cube comes into play when we're dealing with data that exceeds the three dimensions that are commonly found in everyday data analysis. (23)

2.3. Clinical Data repository benefits

A robust clinical data repository can offer a variety of benefits to its users, enhancing collaboration and business operations. Especially in today's dynamic landscape of evolving processing and storage technology, the advantages of consolidating datasets and databases in a central repository are more significant than before. The establishment and utilization of a data repository come with numerous benefits (24).

To begin with, data repositories play a pivotal role in enhancing the precision and correctness of data, leading to quicker and more precise decision-making. Accessing accurate information exactly when it's needed can significantly impact critical business choices and contribute to the prioritization of ongoing projects. Especially when dealing with crucial items that play a role in vital public services, such as GPS or healthcare, having reliable and precise data is of paramount importance. This requirement extends to the realm of space exploration as well, where accurately determining object positions and orientations is essential for interpreting potential object collisions and making sense of sensor data collected from satellites and medical devices. Hower, the question arises is how does a data repository ensure that the data it houses is both of exceptional quality and utmost integrity? According to a study on The Economics of Data Integrity conducted by the Online Computer Library (OCLC), guaranteeing data quality demands specialized expertise in specific subject areas. This is because the diverse needs of different disciplines necessitate tailored approaches to maintaining data integrity. (25).

In today's business landscape, data accuracy and reliability hold immense importance for every modern organization and company. They stand as fundamental pillars, directly influencing the precision and effectiveness of all business operations and decision-making processes. Moreover, data quality and integrity serve as the core focal points within most data security initiatives. The pursuit of these aims involves a diverse array of standards and techniques, ranging from meticulous data requirement gathering to meticulous access control measures, thorough validation of data inputs, elimination of duplicate data entries, and the regular implementation of data backups (26).

A critical aspect of achieving these goals involves ensuring that all data is thoroughly and accurately described in a way that aligns with its intended meaning, as recognized by the owners of the respective datasets. This meticulous description fosters a more precise representation of the data and paves the way for meaningful comparisons across datasets that share common data elements. This, in turn, aids in identifying areas that require data cleansing or enhancements in terms of quality. (17)

Secondary, a centralized data repository contributes to the augmentation of data sharing and the fostering of collaborative efforts. As previously mentioned, the integration of datasets via semantic metadata amplifies data quality and integrity due to multiple stakeholders accessing the same logical framework. This centralization of data representation, driven by semantic consistency, also catalyzes collaboration and the exchange of data. Individuals utilizing the same dataset gain the means to connect with fellow users, ultimately nurturing a comprehensive and data-centric environment (12). The advantages outlined above are just a glimpse into the numerous rationales supporting the adoption of a centrally integrated Data Repository as an optimal solution for addressing the challenges posed by healthcare data. Such a semantic Data Repository holds potential for resolving various prevalent data-related issues, such as discrepancies or inaccuracies in data records. Furthermore, it can generate a "network effect" of value meaning that as more data becomes integrated, the resulting value would extend to a broader spectrum, benefitting healthcare services, operational aspects, and research endeavors alike.(27)

Thirdly, a central data repository serves as a means of safeguarding historical data while also facilitating the assessment of goods and services. The significance of datasets frequently emanates from recent data, such as the latest health device readings that provide essential insights into the current health status of patients. Yet, historical datasets hold their own substantial value, primarily for trend analysis. They aid in examining potential shifts in baseline measurements over time or detecting alterations in data collection methodologies (12).

The integration of data within a centralized repository, coupled with an appropriate storage infrastructure, ensures the seamless retention of historical data. This, in turn, allows for smooth upgrades or modifications of applications that tap into that data. Notably, the preservation of historical data goes beyond mere technical considerations. It leads to the generation of deeper insights, ultimately translating to enhanced economic value. For instance, preserving historical patient data spanning various healthcare facilities empowers improved forecasting and informed decisions for future treatment strategies. (17)

Another advantage lies in centralizing the storage and upkeep of data. The integration of data within a central repository or database empowers users to make changes that are instantly reflected across the organization. This mitigates the challenge of managing multiple sets of data that necessitate continuous synchronization efforts. Moreover, this approach fosters a cohesive workflow across various departments, allowing multiple users to access the same real-time data. Furthermore, a data repository effectively curtails time consumption and eliminates redundancies. Through data integration and collaborative efforts, redundancies are notably reduced, and in many cases, entirely eradicated. The removal of repetitive information translates to quicker review processes and expedited decision-making, culminating in heightened productivity. This heightened level of collaboration across departments or organizations ultimately saves valuable time for decision-makers, endowing the organization with a distinct competitive edge.(28)

Lastly, it contributes to yielding a more substantial Return on Investment (ROI). The initial costs of establishing a data management strategy might appear substantial. However, when contrasted with the expenses linked to poor data quality, erroneous information, and uninformed decision-making, the advantages become evident. Data points to the fact that organizations implementing central data repositories have not only increased their revenue but also made significant cost savings compared to those that haven't made such investments (29).

Considering data as a pivotal organizational asset, the effective management of data quality holds paramount importance for the success and even survival of a business. The manifold benefits of adopting a central data repository result in improved decision-making, heightened productivity, enhanced efficiency, and better-informed employees within the organizational framework. The implementation of a central data repository essentially facilitates the optimization of an organization's business processes, enabling them to function at their peak potential.(28)

2.4. Architecture of Health Data Warehouse

The structure of the national health data warehouse (DW) model is depicted in Figure 3. Health information sourced from various government and private entities, including hospitals, clinics, diagnostic centers, and research facilities, will be gathered. Through the Extract, Transform, Load (ETL) process, this data will be consolidated into a temporary data repository, followed by subsequent steps involving data cleansing, noise reduction, and normalization techniques. Subsequently, the prepared data will be loaded into the data warehouse (DW). This streamlined process enables the execution of Online Analytical Processing (OLAP) queries and data mining operations with ease, making the pathological DW a valuable resource for healthcare analysis (30).

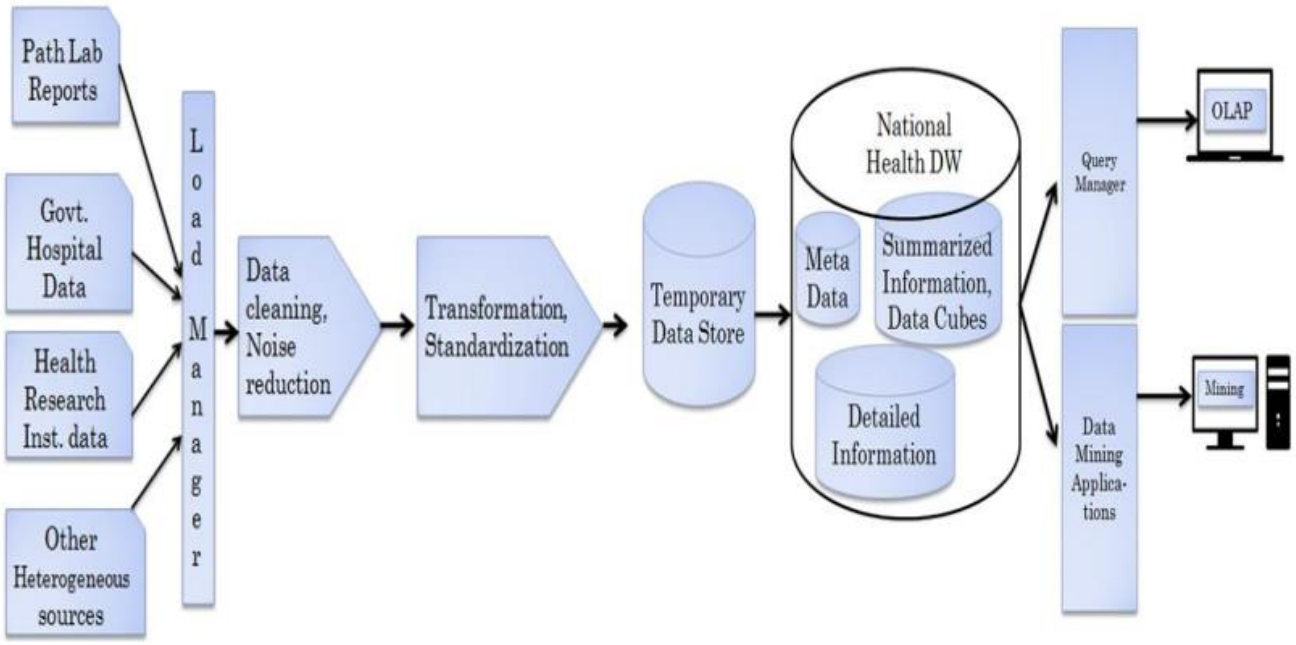

Figure 5.Brief Architecture of Health Data Warehouse

A high-level depiction provides an overview of the workflow within the clinical data warehouse solution, covering operational databases, external data sources, interfaces for Extraction Transformation Loading (ETL), the data warehouse, data mart, and data analysis, as shown in Figure 4. Unlike the conventional approach, which typically begins with data extraction from operational databases and external sources, followed by ETL processes to populate the warehouse, our team recognized several shortcomings associated with this approach when implemented at other institutions. Many of these projects faced a fundamental design challenge where the focus was predominantly on extracting data from sources,

lacking clear use-case scenarios and data mapping requirements. Acknowledging the potential pitfalls of this approach, our team chose to adopt a “backward-in” strategy for the project (30).

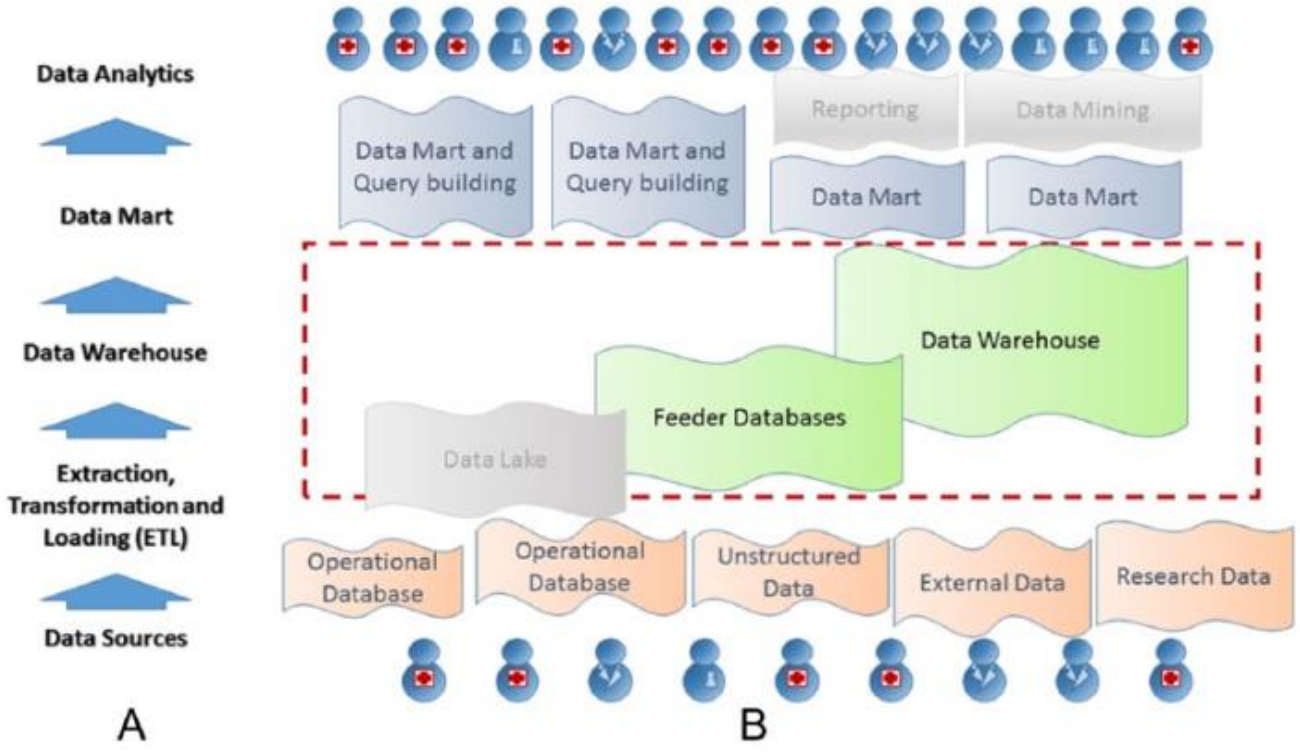

Figure 6. Architecture of the clinical data warehouse project

In the Architecture of the clinical data warehouse project above. (A) Key layers of the data warehouse layout. (B) Components in the implementation. The data lake component as well as further reporting and mining tools have not yet been implemented and are therefore rendered in gray.

As it was suggested by the results study conducted in 2017 entitled Roadmap to a Comprehensive Clinical Data Warehouse for Precision Medicine Applications in Oncology the system should have on The Web image archive applications that provide access adhere to the following requirements: Firstly, the application must ensure proper image visualization via a Web browser, catering to specific display needs for various medical image types. Efficient navigation for whole slide images entails panning and zooming capabilities, while stack support is necessary for computed tomographic images. Secondly, direct URL access to each image should be facilitated, eliminating the need for users to search for patient or timestamp information. Thirdly, URLs can be automatically fetched from the application's back-end database into the clinical database along with complementary image metadata. Fourthly, image access

is restricted by user account and password, meaning that even if an URL is shared, the recipient must possess proper access privileges. Fifthly, the application server is strategically configured to ensure network security within the institution. Lastly, for users without access to Protected Health Information (PHI), image access is tailored to remove sensitive data, potentially necessitating PHI-free image versions to be served or a separate configuration for a PHI-free image server (30).

2.5. Trustworthy Healthcare Data Repositories Requirements

2.5.1 Organizational Infrastructure

At its forefront, a healthcare data repository should establish a well-defined mission and scope. This mission should be unequivocally oriented towards granting access to and safeguarding healthcare-related data. The repository bears the responsibility of being a dedicated steward for digital patient data, guaranteeing that these materials are securely maintained in suitable environments for the required durations. Both service providers and users should be fully aware that the repository's explicit role involves not only the preservation of data but also the ongoing assurance of access to that data. (31).

Secondly, the repository needs to establish a fitting rights model that encompasses data access and usage. This model must be communicated clearly to users, and adherence to it should be vigilantly monitored. This stipulation pertains to the access protocols and licensing guidelines stipulated by the repository itself, as well as any ethical codes prevalent in the healthcare domain for the responsible sharing and utilization of knowledge and information. It's imperative that the repository substantiates its robust controls in alignment with the defined data access criteria, along with furnishing evidence of effective management of pertinent licenses and procedures. (31)

Thirdly, there's the imperative of maintaining Continuity of access. The repository needs a robust plan to guarantee continuous access to its data while safeguarding its integrity. This aspect encompasses the governance framework that pertains to the repository's uninterrupted functioning over extended periods and during unforeseen events. It also entails substantiating the existence of a succession plan – detailing the strategies implemented to ensure ongoing access to and availability of the repository's data holdings, both in the present and down the line.

Reviewers are particularly interested in discerning evidence that signifies proactive preparations to address the inherent risks associated with changing circumstances. This includes not only shifts in the repository's mission and scope but also potential alterations in the operational landscape. (32).

Lastly, there's the pivotal aspect of Confidentiality/Ethics. The repository commits to upholding data creation, curation, access, and usage in alignment with the ethical standards and disciplinary norms of the field. The repository strives to ensure that its practices are in harmony with these ethical considerations to the best of its ability. Crucially, evidence should underline the repository's adeptness in implementing sound protocols for data that carries potential disclosure risks. This includes providing comprehensive guidance for users in handling such sensitive data. This diligence is indispensable to sustain the trust of patients who consent to the storage of their personal and sensitive data within the repository. (31).

2.5.2 Clinical Data Management

Clinical Data integrity, security, and authenticity stand as paramount requirements in effective data management. The healthcare repository is under the obligation to provide compelling evidence of a robust data and metadata management system. This system should be proficient in ensuring the solid integrity and authenticity of data across pivotal stages such as data ingestion, archival storage, and patient data access. This criterion envelops the entire data lifecycle within the repository, highlighting the significance of a comprehensive and seamless strategy for upholding data precision and credibility (25). In order to safeguard the integrity of both data and metadata, deliberate alterations made to them should be diligently documented, along with the rationale behind such changes and the entity responsible for them. Mechanisms must be in place to promptly identify unintended or unauthorized modifications, allowing for the restoration of correct versions of data and metadata. To enhance the level of documentation for all integrity and authenticity procedures, an augmentation in the monitoring of changes – including the logging of individuals effecting modifications – should be incorporated into the repository's roster of initiatives(33). The aspect of authenticity encompasses the reliability of the original deposited data and its provenance. This extends to establishing the connection between the initial data and the disseminated versions, along with the preservation of existing relationships between datasets and/or metadata (31).

In the realm of big data, ensuring security and privacy holds utmost significance. Privacy essentially revolves around safeguarding sensitive information pertaining to personally identifiable healthcare data. Its focus lies in managing and overseeing individual data, entailing the formulation of policies and establishment of authorization prerequisites. These measures ensure that patients' personal information is acquired, shared, and utilized in an ethical manner by authorized entities (12). On the other hand,

security pertains to shielding data from unauthorized access, with certain definitions also encompassing the concepts of integrity and availability. The core emphasis here is on fortifying data against malicious breaches and preventing the theft of data for illicit gains. While security is undeniably critical for data protection, it falls short in fully addressing privacy concerns (25). In today's landscape, it's imperative that healthcare organizations effectively manage and shield personal information. This involves addressing associated risks and fulfilling legal obligations related to the processing of personal data, especially given the intricate web of data protection legislation that continues to expand. (34).

Moving on to the next necessity, researcher encounter the appraisal aspect. In this context, the repository's role is to accept data and metadata adhering to established standards, a practice aimed at guaranteeing the pertinence and comprehensibility of data for its users. This appraisal function bears pivotal significance, as it involves the critical evaluation of data to ascertain if they meet the specified criteria for inclusion, and subsequently, ensuring proper management for their enduring preservation. Continuous appraisal and periodic review are essential to ascertain that data remain pertinent and comprehensible to the intended user community (35). In addition, there's the requirement of a preservation plan. The repository undertakes the responsibility of safeguarding data for the long term, carrying out this role in a methodical and well-documented manner. This plan encompasses strategies to ensure the continued protection and accessibility of data over time, aligning with the repository's commitment to responsible data management. (26).

Ensuring the quality of data is an essential aspect of managing data. When it comes to healthcare data, the data repository should possess the necessary expertise to handle both the technical aspects of data and metadata quality. This expertise ensures that healthcare professionals and other users can rely on the data to make informed decisions related to quality. The repository's role includes assessing the completeness and excellence of both the data and metadata. Furthermore, the repository supports the ongoing usability of the data, making sure that the relevant metadata are available to aid in understanding and using the data in healthcare services. A crucial aspect is that repositories must maintain the data's effectiveness and comprehensibility in the face of changing technologies and the evolving knowledge base of the intended users. This requirement focuses on the steps taken to guarantee that the data remains reusable in the long term (12).

2.5.3 Health Data Repository Technology

Around the mid-2000s, there was a significant transformation in how we manage large volumes of data. This marked the time when conventional systems designed to handle substantial data loads started to become parallelized. Before this change, the approach for dealing with what we consider "larger" data involved upgrading through 'vertical' scaling, where we relied on faster processors and increased storage capacity (15). However, with the emergence of the big data paradigm, there was a notable shift towards distributing data 'horizontally' across numerous nodes. This new approach allowed for scalability, efficiency, and cost-effectiveness in processing. This transition to parallelization in managing data-intensive tasks can be likened to a similar shift that occurred years ago in compute-intensive systems, a field known as High-Performance Computing. The realm of big data systems can be classified according to three main components: the infrastructure supporting them, the platforms used for storage, and the frameworks utilized for processing. (36).

When it comes to the infrastructure of big data systems, there are a few different approaches. One option is utilizing cloud resources, while another is setting up systems on physical hardware or bare-metal servers. Cloud infrastructures, like Amazon Web Services Redshift, bring extra tools to the table, but many tools can work effectively in both cloud environments and on on-premises resources. This adaptability allows for a flexible deployment based on specific needs (32). The evolution of storage platforms started with the expansion of file systems to handle larger block sizes, and this was notably seen through the introduction of the Hadoop Distributed File System. Following this, non-relational databases, also known as NoSQL databases, came into play. These databases allowed data to be distributed across multiple data nodes for parallel processing. While they sacrificed some of the properties typical of relational databases, like structured data organization, they gained speed by enabling independent processing across various data nodes. The initial types of NoSQL databases included key-value, document, big table, and graph databases. In time, further developments took place to reintroduce some relational database features into distributed data systems. These advancements coined the term "NewSQL" to describe systems that aim to restore relational database properties to these distributed environments (37).

The shift in data processing was initiated by the introduction of the MapReduce (M-R) technique. With MapReduce, a single data processing query was distributed to numerous individual data nodes, and the outcomes were collected back to a central compute node. However, the execution of tasks within this

framework had certain inefficiencies. This prompted the development of MapReduce 2.0, also known as YARN, which enhanced the management of cluster resources (38). In the years that followed, additional analytic frameworks like Spark were introduced. These frameworks aimed to shield data science applications from the intricacies of lower-level implementation, providing a higher level of abstraction and simplifying the process of data analysis. (1).

The design of a big data system is contingent on the attributes of the dataset itself, which encompass its size, speed of accumulation, diversity, and the extent to which its features vary. The process of analytics follows a sequence that involves gathering data, refining it (by cleaning and organizing), analysing it, and subsequently taking informed actions based on the insights derived. (17). In conventional data warehousing, data is initially amassed and temporarily stored, then subjected to curation (achieved through Extract, Transform, Load processes), and finally, the cleaned data is stored within a warehouse. Following this, the data is analyzed, prepared for actionable insights, and subsequently acted upon by the organization. However, in the realm of high-volume data systems, data is stored as it arrives. The sheer amount of data makes moving it impractical, so cleansing and analytics occur dynamically in tandem. For data with high velocity, previously categorized as data streaming, analysis is conducted in real-time, often utilizing in-memory systems. After these analytic or alerting processes, some summarized data may be retained (32).

Incorporating variety data introduces the complexity of integrating very distinct datasets, sometimes from sources outside the organization's ownership. Variability, on the other hand, relates to datasets with changing attributes. Architecturally, this implies that the analytics system might require varying resources, depending on the circumstances. This type of fluctuating data processing, often referred to as "bursty," can be efficiently managed through cloud systems. These platforms allow for the dynamic addition or removal of computational resources as required, thus optimizing costs (38).

When considering the bigger picture, the selection of tools and methodologies for infrastructure, storage platforms, and processing collectively forms an enterprise's data repository. The adoption of NoSQL systems within the storage platform has given rise to the concept of a "data lake." In a data lake, diverse datasets are stored within different NoSQL databases. Subsequently, the process involves integrating and processing the specific data required for a particular application. The notion of a data lake often signifies that an organization is adopting a centralized approach to manage and oversee all of its data. (39).

2.5. Data repository risks and limitation

The multitude of terms such as Digital health, e-health, and telemedicine underscores the ongoing scientific and technical revolution catalysed by rapid health data processing and artificial intelligence (AI). This transformative phase is leading to significant changes within health systems. Beyond mere data transmission and supporting medical and administrative management in hospitals, Big Data offers unprecedented opportunities for monitoring population health, medical decision-making, and risk assessment. The convergence of AI and Big Data is expected to profoundly reshape medical practices, healthcare delivery, and health research, finding applications from fundamental research to diagnostic, predictive, and therapeutic decision support tools. The advancement of data science and bioinformatics is indispensable in modelling the impact of various anomalies on individual organisms.

In other words, the array of terms such as Digital health, e-health, and telemedicine highlights the ongoing revolution enabled by rapid health data processing and artificial intelligence (AI), bringing about significant changes in health systems. Beyond facilitating data transmission and aiding medical and administrative management, Big Data offers unprecedented opportunities for population health monitoring, medical decision support, and risk assessment. The amalgamation of AI and Big Data is poised to deeply reshape medical practices, healthcare delivery, and health research, finding applications from basic research to diagnostic, predictive, and therapeutic decision support tools. The progression of data science and bioinformatics is crucial for effectively modelling the impact of various anomalies on individual organisms (40).

Moreover, there's a notable concern surrounding security and privacy risks. The breach of a patient's privacy becomes a reality if attackers can deduce the data owner's identity through content analysis. The stakes are considerably higher when all data is consolidated in a single location; an attack on such a repository puts the entire country's healthcare data in jeopardy. Unauthorized individuals could potentially gain access to critical information and exploit it in unauthorized ways (24). Another dimension of risk pertains to the mishandling of patient data, which poses security and privacy threats. This can manifest in various ways, including security breaches that may expose health information to malicious cyberattacks. These attacks can result in the de-anonymization of patients' personal data and diagnoses, the commercialization of data, and the illicit sharing of patient information for unlawful purposes. When confidential patient data falls into the wrong hands, it can lead to undesirable and adverse consequences (37).

Another potential risk concerns the limitation of storage capacity, which might be constrained. While a data repository offers the advantage of securely housing an extensive compilation of patient medical records, the practicality of uploading streamed data onto it is hindered by the sheer volume of data that accumulates over time. Furthermore, reliance on cloud servers occasionally introduces operational vulnerabilities that can result in data unavailability due to potential operational failures. (24).

Lastly, the quality of performance stands as another factor of consideration. This encompasses processing speed, denoting the time taken for uploading, downloading, and processing patient health data. Interoperability is another facet, denoting a storage medium's capability to seamlessly exchange data across diverse systems and software. Furthermore, data transparency plays a role, indicating a storage medium's aptitude to ensure accuracy, the authenticity of data sources, and the ease of data access and utilization, regardless of origin or location. Given the substantial volume of data it holds and the incoming traffic it experiences over time, there's a possibility that the performance of a data repository might gradually slow down. This scenario can potentially impact the efficiency of healthcare services delivery. (39).

2.7 Research gaps

Despite the growing recognition of the potential benefits of implementing a national health data repository in Rwanda to enhance patient data accessibility and healthcare service delivery, there remains a dearth of empirical studies that specifically focus on the challenges and strategies associated with the integration of such a repository within the context of referral and district hospitals. While existing literature acknowledges the advantages of centralized health data management, there is limited research that delves into the nuanced operational, technical, and organizational barriers that might arise in the implementation process at these specific levels of the healthcare system.

The extensive discourse on the transformative potential of digital health, e-health, telemedicine, Big Data, and artificial intelligence (AI) in revolutionizing healthcare systems, there remains a need for comprehensive studies that holistically explore the challenges, ethical considerations, and practical implementation strategies that arise from their integration. While numerous studies highlight the benefits of these technologies, there is a relative scarcity of research that delves into the potential drawbacks, unintended consequences, and barriers that healthcare organizations and practitioners might encounter when adopting these innovations. Additionally, the literature tends to focus on the technological aspects, leaving room for in-depth investigations into the impact of these advancements on patient-provider

relationships, the quality of care, and the potential social implications within diverse healthcare contexts. Therefore, further research is required to provide a well-rounded understanding of the complexities surrounding the adoption and utilization of these technologies, addressing both their potential benefits and the challenges they introduce across medical practices, patient care, ethics, and societal dynamics.

Addressing this gap is crucial as referral and district hospitals play a pivotal role in delivering healthcare services to diverse populations, often facing unique constraints and resource limitations. Understanding how the establishment of a national health data repository would impact these hospitals' daily operations, clinical decision-making, and patient outcomes is essential for tailoring effective implementation strategies. Furthermore, exploring the potential for interoperability between these hospitals and the repository, along with any potential resistance from healthcare practitioners, administrative staff, or patients, would provide valuable insights for policymakers and healthcare managers seeking to ensure a smooth transition to this integrated data management system.

Therefore, research that investigates the specific challenges, opportunities, and potential mitigations related to implementing a national health data repository within the referral and district hospital setting in Rwanda would contribute significantly to the current literature and offer practical insights for successful implementation.

CHAPTER 3: RESEARCH METHODOLOGY

3.1 Study Approach

The study was conducted using quantitative because it was respectively included the collection and evaluation of numerical data to identify a pattern within numbers.

3.2 Study Design

The study design was cross-sectional, and it conducted investigations to answer specific clinical questions or solve practice-related problems focusing on health data management, data integration, reporting, and knowledge discovery.

3.3 Study setting

This research was carried out across two nationally recognized referral hospitals located in Kigali, namely King Faisal Hospital located in Kacyiru, Gasabo District, Ndera Neuropsychiatric Teaching Hospital located in Ndera Sector, Gasabo District. One Level two teaching hospital (Nyamata Level Two Teaching Hospital) located at Nyamata, Bugesera District, and two district hospitals namely Gatonde District Hospital located at Mugunga Sector, Gakenke District, and Nyarugenge District Hospital located at Nyamirambo sector, Nyarugenge district. These hospitals have a primary objective of offering advanced medical care to the public. They provide specialized services in various fields like Gynecology and Obstetrics, Internal Medicine, Clinical Specialties, Pediatrics, Dentistry, Surgery, and Mental Health. The researcher chooses these hospitals are some of the pivotal in Rwanda, catering to complex cases that are referred from different districts seeking specialized treatment. These hospitals are staffed by a diverse group of professionals in the clinical domain, including specialists from both Rwanda and other countries, general practitioners, allied healthcare professionals, and registered nurses.

3.4 Study population

The target population of this study mainly included health care providers of the mentioned hospitals in specialized care like internal medicine, surgery, gynecology and obstetrics, clinical specialties and mental health, and staff from other necessary units connected to health data generation and use, health data management and data integration include ICT staff and data managers.

3.5 Study sample

In this study, two distinct sample groups were considered. The primary group under investigation comprised healthcare personnel, including both doctors and nurses. These individuals are responsible

for documenting patients' data and utilizing it for making informed decisions during the care process. The secondary group, on the other hand, comprised Information and Communication Technology (ICT) staff and data directors. The study's overall sample encompassed a total of 110 healthcare providers, along with 10 ICT and data management professionals selected from the previously mentioned hospitals convenience sampling methods. Hence, the cumulative study sample comprised 120 participants.

3.6 Sampling strategy

In the sampling approach, the sample size was determined using convenience sampling as a non-probability sampling method where units were selected for inclusion in the sample due to their availability and willing to participate in the research where the duration was two months starting from April 1st, 2023, to June 30, 2023. During that period, 120 participants were selected from the selected hospitals. They were invited to provide responses to semi-structured questionnaires. Finally, to complete our research, researcher also engaged participants who were conveniently available for the study.

3.7 Inclusion criteria

For this study, the criteria for inclusion revolved around three distinct participant categories: medical doctors, ICT staff, and data managers. Within the medical staff category, individuals specialized in healthcare fields such as radiology, surgery, internal medicine, and clinical specialties were eligible. Participation was contingent upon their willingness to sign the consent forms voluntarily and engage in the study. In addition, ICT staff (head of ICT the department and Data Manager) who completed the probation period and signed consent forms from each selected hospital.

3.8 Exclusion criteria

The staff members who have not attained six months of experience were considered to be in the probation period and were excluded from the study. In addition, staff members who refused to sign the consent form were also excluded.

3.9 Data collection

In this study, different groups of respondents are identified, but each gets a single type of questionnaire, so there is only one data collection method. Research used interviewer-administered semi-structured questionnaires to all respondents, either in English or Kinyarwanda, depending on respondent language preference. The data collection was done as follows:

- Researcher inform the participants about the plans for the central repository through the art of presentation of 15 minutes, the presentation was neutral just explaining how it will work not

mentioning any advantages or risks, and then distribute questionnaires to be answered in the following 20min to manage the time since healthcare providers may have a lot of work to do.

- Semi-structured questionnaires for respondents to get insights on how current health data flow affects Patients in receiving healthcare services, clinical benefits and challenges of the current health information exchange and their needs for a national health data repository to facilitate single patient records and advantages of developing a national health data repository.

The presentation was conducted based on informal appointments (face-to-face) this is for medical staff and ICT personnel. The data collection was started after being authorized by the UR-College of Medicine and Health Sciences (CMHS) Institutional Review Board (IRB). Thereafter, researcher write a letter to the managing director of hospitals requesting permission for data collection. After getting the response from the hospital managing director as a permit for data collection, the researcher scheduled the presentation and explained the purpose of this research.

At the time of data collection, the participants who were available explained the study briefly and asked to volunteer and consent to participate in the study. After signing the consent, researcher gave them the questionnaire which was grouped into 3 sections. The second focused on examining the level of need and support among participants for the implementation of a national healthcare data repository and the adoption of a Single Patient Record system, which is our first objective. In addition, to address the second and third objectives researcher asked questions concerning the Clinical Benefits, barriers, and the best approaches that can be used for the effective implementation of a national healthcare data repository.

Finally, the study covered a period from November 2022 to Aust 2023 during which sufficient data was collected, analyzed, and interpreted.

3.10 Data management and Analysis

The questionnaires were meticulously reviewed to ensure their completeness. To analyze the quantitative data, researcher utilized SPSS software version 26. Researcher employed cross-tabulation techniques to examine potential relationships between variables. To facilitate this, the data were initially entered into an Excel worksheet, which streamlined their organization. Subsequently, the data was imported into SPSS version 26 for analysis. A significance level of $p \leq 0.05$ was chosen. To provide an overview of the data, researcher utilized frequencies as part of the descriptive statistics. In this study, the independent variable was the implementation of a national health data repository in Rwanda. On the other hand, the dependent variables were patient data accessibility and healthcare service delivery within Rwanda.

3.11 Problems and limitations of the study

The major problem for conducting this study was availability of the healthcare providers. This study required a greater deal of time for data collection and the availability of respondents. However, the target respondents of this study have great responsibility which were a great limitation to the researcher in getting enough time with them during data collection. Another problem for conducting this study was the shortage of funds, due to time and financial constraints, we were limiting the extensity of our research to two Kigali Based national referral and one teaching hospitals and two district hospitals. The replication of the study at another hospital would enable the better generalization of the findings revealed interesting.

3.12 Ethical Consideration

The study proposal was submitted to the University of Rwanda, College of Medicine, and Health Sciences Institution Review Board (IRB) for review and ethical clearance approval. Then permission to carry out the study in the hospital were obtained from each clinical research department from of hospitals after reviewing the proposal. The respondents were informed about the study objectives before given the questionnaire and the data collected remains confidential and used only for the stated research purposes. Participation in this research was entirely voluntary and there were no punishments for those who decided not to participate in the study. The respondents were requested to sign an informed consent form. To guarantee confidentiality, the respondents were reassured that the questionnaire forms will not show their names but will be identified by the codes. The authors mentioned in this study were acknowledged within the references. The results obtained from the study allowed the researcher to interpret the data, formulate conclusions, and present recommendations in a generalized depersonalized manner. Finally, no patient-related data at all were collected or processed in the study, only the personal opinions of informed healthcare professionals. Therefore, there is no delicate ethical problem linked to this study.

CHAPTER 4. RESULTS PRESENTATION

In this chapter, researcher delve into a comprehensive discussion of the key findings derived from the data analysis of our study. The results provide valuable insights into various aspects related to participants' perspectives on healthcare data management, access, and the potential implementation of a national health data repository. The research findings in this context were organized to investigate whether it's needed, practical and possible to implement a national health data repository in Rwanda. The primary focus was to enhancing patient data accessibility and healthcare service delivery, with a specific case study centered around referral and district hospitals. The study systematically analyzes the resources required for implementation, potential challenges, and benefits associated with the repository.

By conducting a comprehensive study and examining the practical implications at different levels of the healthcare system, this research aimed to understand how a centralized health data repository could revolutionize healthcare in Rwanda. The potential advantages include improved data exchange, better decision-making, and increased opportunities for research and innovation. However, challenges such as data privacy and financial implications also need to be considered. Ultimately, the findings of this study could have a significant impact on healthcare and patient outcomes in the country.

4.1 Background Information of Respondents

In this section of the questionnaire, participants were asked questions related to their demographic data such as background education or gender.

4.1.1 Gender of the participants

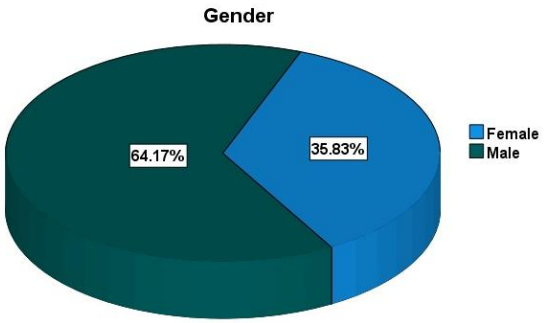

Figure.7 Research participants

The results underscore an unbalanced gender representation among the participants. Specifically, the female respondents accounted for 43 individuals, comprising 35.83% of the total sample. In contrast, the male respondents constituted the majority with 77 individuals, constituting 64.17% of the entire sample. This dataset holds significant implications as it furnishes valuable insights into the gender makeup within the surveyed group. The predominance of males over females in this subset prompts intriguing inquiries and invites potential further explorations into the underlying factors or biases that may influence such an observed distribution. While it's imperative to acknowledge that this survey's scope might not fully encapsulate the broader population, given its reliance on a specific sample size, the outcomes nevertheless offer a window into the gender dynamics within this segment. These insights could prove pertinent for specific contexts and decision-making processes, shedding light on potential gender-related considerations. In conclusion, this data highlights the importance of examining gender distribution in various settings to gain insights into potential disparities and disparities. The observed gender disparity in this sample could prompt further investigations and discussions on gender-related issues, which may ultimately contribute to more inclusive and equitable policies and practices.

4.1.2 Age of the participants

The graph presented below visually depicts the diverse age distribution observed among the research participants.

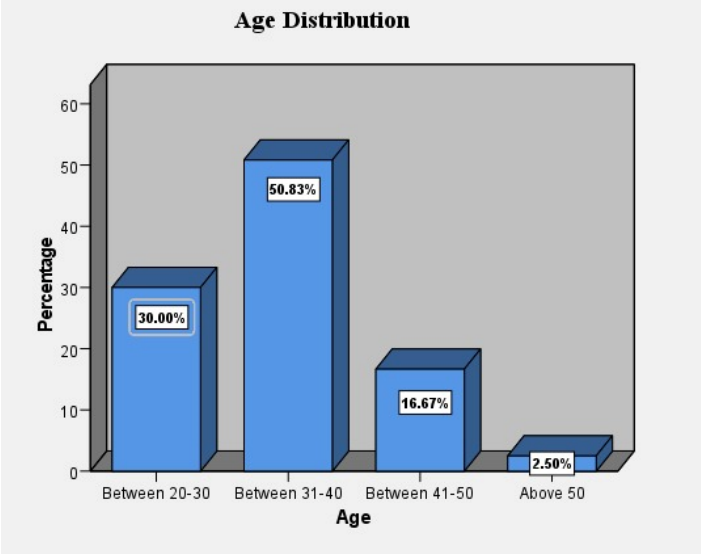

Figure 8: Age of the participants

The findings from the research participants provide valuable insights into the development of a National Healthcare Data Repository and Single Patient Record, particularly concerning the age distribution of the individuals involved. The data analysis results reveal a diverse age distribution among the participants. The majority of respondents fall within the age range of 31 to 40 years, constituting 50.8% of the sample. Participants aged between 20 and 30 years represent 30% of the group, highlighting a significant presence of young adults. Moreover, individuals aged between 41 and 50 years account for 16.7% of the participants, indicating a relatively balanced representation across middle-aged individuals. Notably, those above 50 years constitute a smaller portion at 2.5%. These findings underscore the importance of accounting for various age groups in the interpretation of the study's outcomes and implications, acknowledging the differing perspectives and experiences that participants from different age brackets may bring to the research.

This diverse age representation highlights the importance of accommodating various generational perspectives and healthcare needs in the development of the National Healthcare Data Repository and Single Patient Record. Such comprehensive insights ensure that the repository is well-equipped to cater to the distinct requirements and expectations of individuals across different age groups, leading to a more inclusive and effective healthcare system overall.

4.1.3 Participants Professional Distribution

The distribution of professional backgrounds among the surveyed individuals reveals an interesting pattern. Healthcare Takes Center Stage While IT/Data Management Follows as shown in the following graph.

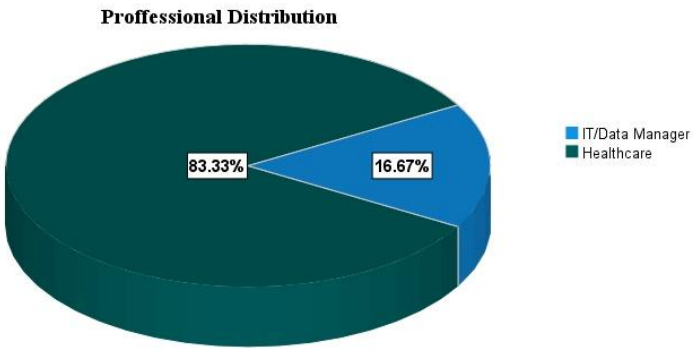

Figure 9.Participants Professional

The above graph portrays a distinct professional composition among the research participants. Notably, the healthcare sector dominates the sample, encompassing 83.33% of respondents. This significant representation underscores the study's potential relevance and impact within the healthcare domain. Moreover, a smaller yet noteworthy proportion of participants, 16.67%, hail from the IT and data management sector. This subset likely contributes diverse technical perspectives and expertise, enhancing the research's multidisciplinary character. The study's findings should therefore be interpreted with an awareness of the concentrated healthcare presence while also recognizing the valuable input from IT and data management professionals, ensuring a comprehensive and balanced analysis of the results.

4.1.4 Participants level of education

This section presents the ensuing result table that elucidates the education profile of our research participants. The table showcases the distribution of education levels among the respondents, shedding light on the diverse academic backgrounds that enrich our study.

Table 2.Education level

Education level	Frequency	Percentage
Bachelor's degree	77	64.2%
Master's degree	34	28.3%
Doctorate (PhD)	9	7.5%
Total	**120**	**100%**

The data analysis in the table underscores the prevalence of participants holding bachelor's degrees, constituting 64.2% of the sample. This majority representation indicates a strong engagement from individuals with undergraduate education, suggesting a broad-based interest in the research subject. Notably, those with master's degrees comprise 28.3% of the participants, contributing a substantial segment of participants with advanced academic training and potentially deeper subject knowledge. Furthermore, the presence of individuals with Doctorate (PhD) qualifications at 7.5% reflects a smaller yet significant cohort of highly specialized and experienced participants, likely offering nuanced insights into the study's domain. The diverse educational backgrounds of our participants lend credibility and a comprehensive perspective to the research outcomes, promoting a more holistic interpretation of the findings.

4.1.5 Experience of the respondents

The graphical representation that follows, illustrating the distribution of research participants' years of experience. The graph provides an insightful visualization of the diverse experience levels encompassed within our study cohort.

Table 3.Experience of the participants

Years of experience	Frequency	Percent
Below 1 Year	2	1.7%
Between 1-5 Years	40	33.3%
Between 6-10 Years	47	39.2%
Between 11-15 Years	16	13.3%
Above	15	12.5%
Total	**120**	**100%**

The distribution of years of experience among the participants reveals interesting insights. The data analysis underscores a balanced spread of experience across the participants. Notably, individuals with 6 to 10 years of experience constitute the largest segment at 39.2%, reflecting a significant group of participants with a substantial duration of involvement in their respective fields. Moreover, those with 1 to 5 years of experience make up 33.3% of the sample, demonstrating a notable presence of early- to mid-career professionals. Participants with 11 to 15 years of experience contribute 13.3%, showcasing a smaller yet valuable contingent of seasoned professionals. Impressively, individuals with over 15 years of experience represent 12.5% of the participants, signifying a cohort with rich expertise. This diversified distribution across experience levels enriches the study by encompassing perspectives from various stages of professional development, ultimately enhancing the depth and breadth of the research insights.

4.2 The needs and level of support among participants for the implementation of a national healthcare data repository

4.2.1 Challenges in accessing patients' data referred from another health facility.

In this section researcher introduce the accompanying graphical representation, which elucidates the outcome of our analysis regarding *participants' experiences in accessing data of patients referred from other health facilities where* researcher *have asked "Have you ever faced any challenges accessing patients' data referred from another health facility?"*. The graph serves to provide a visual overview about if participant have encountered the challenges while accessing patients' data referred from another health facility.

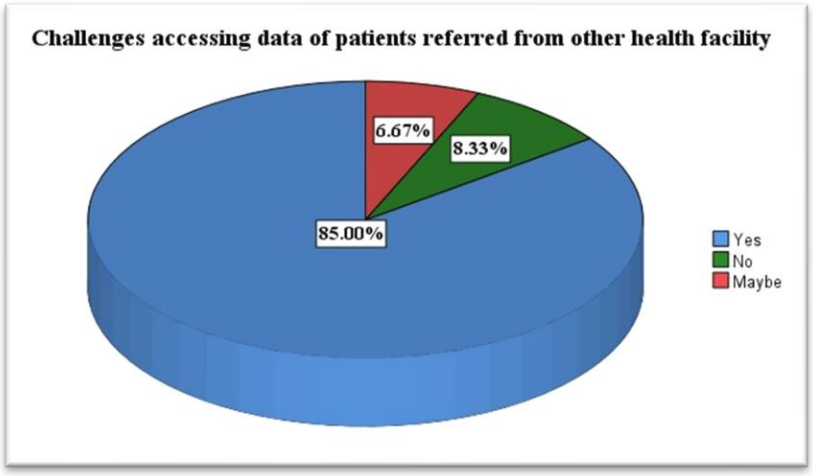

Figure 10. People who have faced challenges in accessing patients' data.

The data analysis brings to light that a substantial majority of participants, accounting for 85%, reported facing challenges when attempting to access data of patients referred from other health facilities. In contrast, a minority of participants, comprising 8.3%, indicated not encountering any challenges in this regard. An additional 6.7% of participants responded with uncertainty, marking a "Maybe" response. These findings accentuate the prevalent difficulties that healthcare professionals encounter when trying to access patient data across different health facilities. The pronounced presence of challenges signifies an area that may warrant further investigation and potential interventions to streamline data accessibility, ultimately contributing to enhanced patient care and information sharing among healthcare entities.

To understand patient data accessibility impact on care delivery researcher, go deeper and ask the following *"How strongly do you agree with the following statement: 'Accessing patients' data referred from another health facility is a challenge to care service delivery since we don't have a holistic view of patients' data.'"* The following graph provides a clear visualization of participants' sentiments regarding the perceived challenge of accessing patients' data referred from other health facilities and its impact on care service delivery.

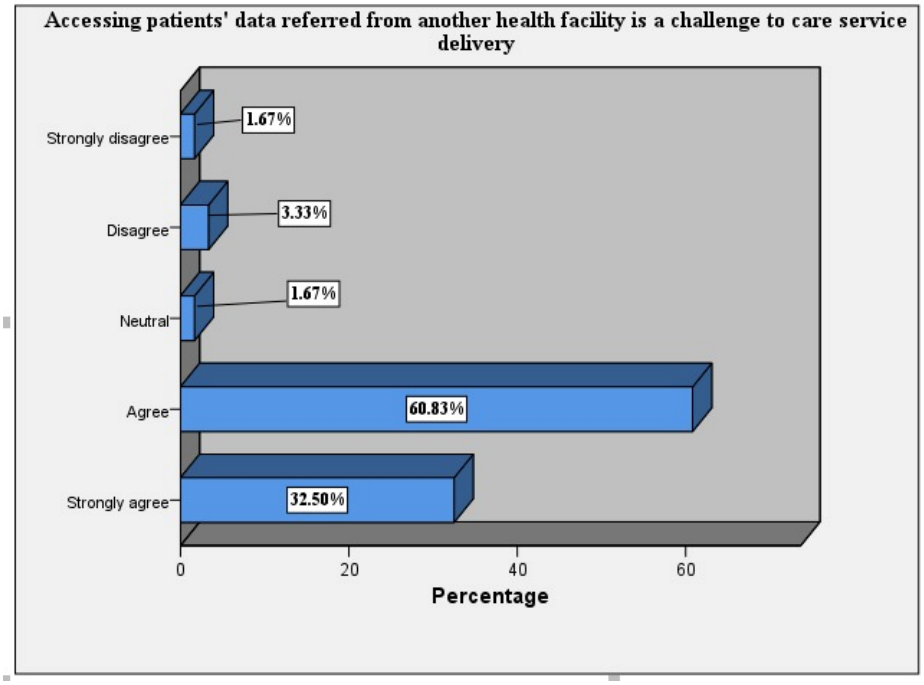

Figure 11. Access patients' data referred from other health facility is a challenge to care service delivery.

The data analysis results unveil a range of opinions among the participants. A significant proportion, constituting 32.5%, "Strongly agree" with the statement, emphatically recognizing the challenge of accessing referred patients' data and acknowledging its detrimental effect on care service delivery due to the absence of a holistic patient data view. Furthermore, 60.8% of participants expressed "Agree," attesting to the challenge posed by the lack of a comprehensive patient data overview from other health facilities. A smaller percentage, 1.7%, indicated a "Neutral" stance, signifying a middle-ground response. A mere 3.3% of participants responded with "Disagree," suggesting that they do not perceive accessing referred patients' data as a substantial challenge to care service delivery. Likewise, another 1.7% "Strongly Disagreed," representing a minority with a strong contrary viewpoint. These outcomes underscore the widespread acknowledgment among participants of the complexities tied to accessing patient data from other health facilities, impacting the seamless delivery of care services. This insight emphasizes the importance of developing solutions to address this challenge and ensure effective communication and data exchange across healthcare entities to support comprehensive patient care.

4.2.2 Satisfaction with your current patient’s data access experience

Here researcher present the accompanying graph, which visually conveys the outcome of our analysis pertaining to participants' satisfaction levels with their current patient data access experience where

researcher asked, “How satisfied are you with your current patients’ data access experience?”. The graph offers a succinct representation of the diverse range of sentiments expressed by the participants.

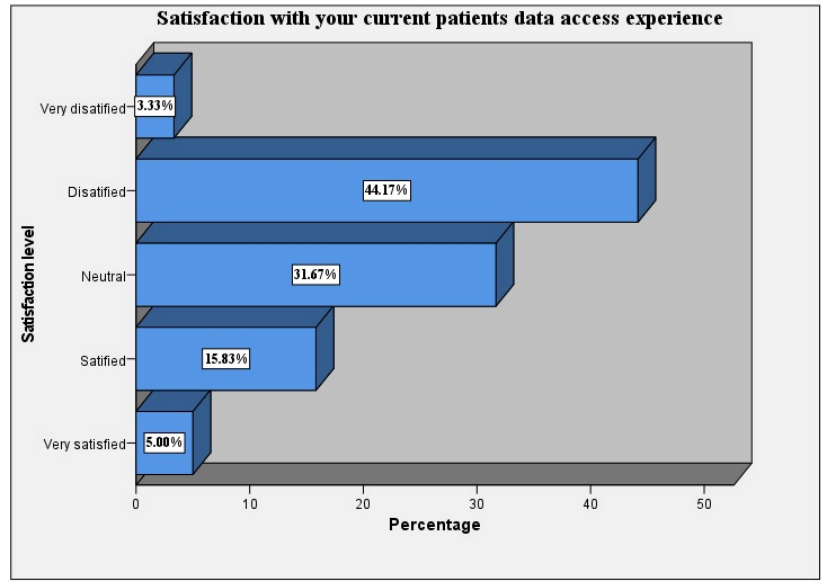

Figure 12. Satisfaction with your current patient’s data access experience

The data analysis results indicate that participants' opinions regarding their patient data access experience are distributed across various levels of satisfaction. A minority of participants, constituting 5%, reported being "Very satisfied" with their current experience. Furthermore, 15.8% of participants expressed being "Satisfied," while a larger portion, accounting for 31.7%, felt "Neutral" about their experience. Notably, a significant portion of participants, representing 44.2%, reported being "Dissatisfied" with their current patient data access experience. Additionally, a smaller subset of participants, at 3.3%, stated that they were "Very dissatisfied." These findings underline the diversity of perspectives among the participants, with a notable proportion expressing dissatisfaction with the existing patient data access process. This highlights a crucial area for improvement, emphasizing the importance of addressing participants' concerns to enhance their data access experience, potentially leading to more effective and efficient patient care and management.

4.2.3 Holistic view of patients’ historical data with current clinical information system

This section introduces the graph that visually presents our analysis of participants' responses to the question: "How strongly do you agree with the following statement: 'The current clinical information system allows me to have a holistic view of patients’ historical data and helps me in decision making to

ensure quality service delivery.'" The graph offers a concise depiction of the participants' sentiments regarding the integration of patients' historical data with the current clinical information system.

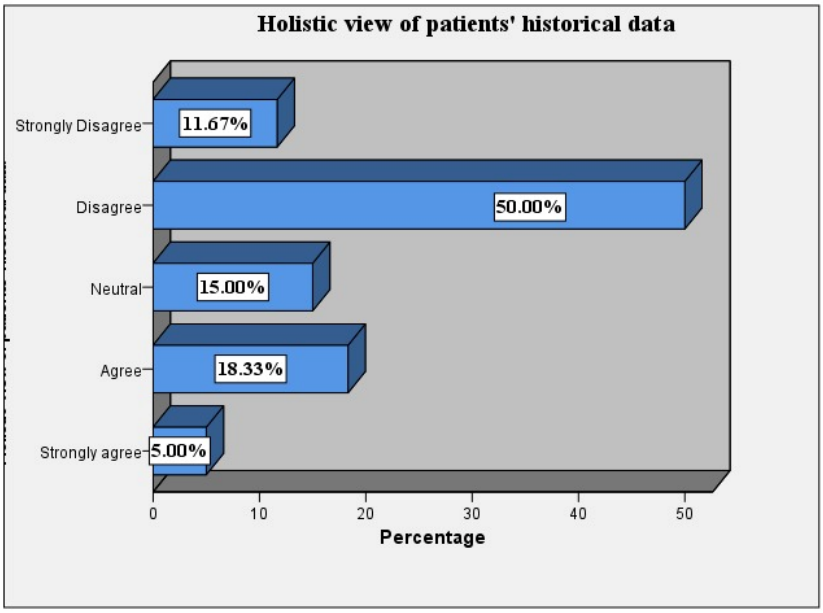

Figure 13. Holistic view of patients' historical data with current clinical information system

The data analysis reveals a spectrum of opinions among the participants. A minority of respondents, constituting 5%, "Strongly agree" with the statement, affirming that the current clinical information system effectively provides them with a holistic view of patients' historical data, thereby supporting their decision-making for quality service delivery. Additionally, 18.3% of participants expressed "Agree," indicating a positive inclination towards the system's capacity to facilitate a comprehensive understanding of patients' historical data. A segment of participants, at 15%, remained "Neutral" about the statement, suggesting a middle-ground perspective. In contrast, 50% of participants expressed disagreement, signifying that they do not perceive the current clinical information system to offer a holistic view of patients' historical data to support their decision-making for quality service delivery. A smaller subset, accounting for 11.7%, "Strongly Disagreed," indicating a strong reservation about the system's ability to meet their expectations in this regard. These outcomes highlight the varied perceptions of the participants regarding the clinical information system's effectiveness in offering a comprehensive perspective on patients' historical data, underscoring an important area for potential enhancement and alignment with healthcare professionals' needs for informed decision-making.

4.2.4 Support level for the idea of developing a national health data repository.

This section includes the graph that visually illustrates our analysis of participants' responses regarding their support for the notion of developing a national health data repository. The graph serves as a visual aid to comprehend participants' inclinations toward the idea of establishing a national health data repository.

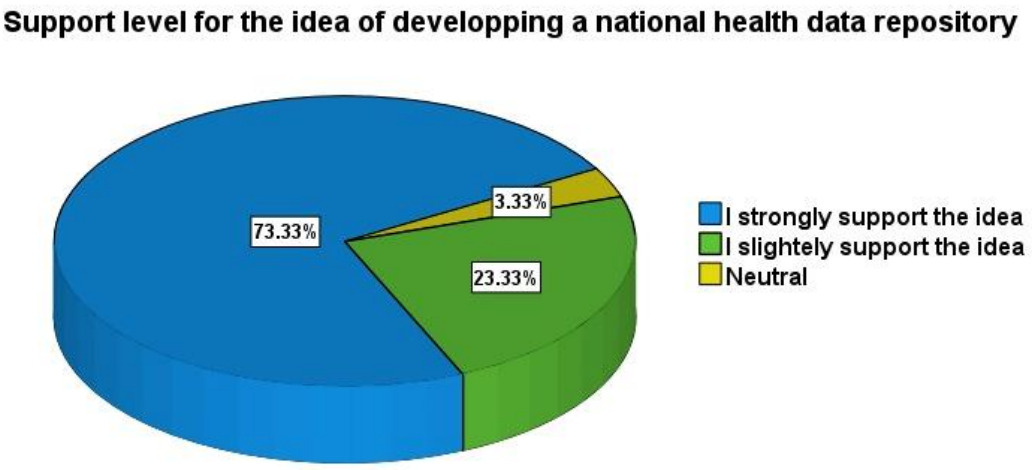

Figure 14. Support level for the idea of developing a national health data repository.

The data analysis outcomes reveal a definitive pattern in participants' attitudes. A significant majority, constituting 73.3%, "Strongly support" the idea of implementing a national health data repository. This substantial endorsement underscores a robust backing for the concept among the participants. Additionally, 23.3% of participants expressed a slightly milder form of support, marking the "Slightly support" category. A minor percentage, at 3.3%, maintained a "Neutral" position on the matter. Notably, there were no responses indicating non-support for the idea ("Slightly I don’t support the idea" or "Strongly I don’t support the idea"). These findings depict a generally favourable disposition towards the establishment of a national health data repository among the surveyed participants. The overwhelming support expressed in both categories highlights the potential perceived benefits of such a repository for healthcare data integration and comprehensive patient care, affirming its relevance and potential utility within the healthcare domain.

4.2.5. Correlation and association

The correlation analysis was conducted to examine the relationship between "Years of experience" and the "Need for a national health data repository and single patient's records." The results are presented in the table below:

Table 4. Relationship between "Years of experience" and the "Need for a national health data repository.

Correlations			
		Years of experience	Need for national health data repository and single patient' records
Years of experience	Pearson Correlation	1	.335**
	Sig. (2-tailed)		.000
	N	120	120
Need for national health data repository and single patient' records	Pearson Correlation	.335**	1
	Sig. (2-tailed)	.000	
	N	120	120
**. Correlation is significant at the 0.01 level (2-tailed).			

The above table serves as a tabular representation of the correlation analysis conducted to discern any potential connection between participants' years of experience and their perceived need for a national health data repository and single patient's records. The results of the analysis reveal a noteworthy correlation between these two variables. The Pearson correlation coefficient, denoted by $r = 0.335$, reflects a positive and statistically significant relationship between participants' years of experience and their perceived need for a national health data repository and single patient's records ($p < 0.01$). This suggests that as participants' years of experience increase, their inclination towards recognizing the importance of a comprehensive health data repository and integrated patient records also tends to rise. The significant correlation underscores the potential impact of professional experience on professionals' understanding of the value and necessity of enhanced health data systems. This finding could have implications for policy discussions, training initiatives, and implementation strategies aimed at

improving healthcare data management and patient care through the lens of practitioners' varied experiences.

4.3 Clinical Benefits of implementing a National Health Data Repository

4.3.1 Participants' Level of Support for Repository as a Solution to patient data accessibility.

The following graph visually presents our analysis of participants' responses regarding their level of support for the statement: "National health data repository and the single patient record would be a solution to patient data accessibility and decision support in healthcare service delivery." It offers a clear visualization of participants' inclinations towards the idea of a national health data repository as a solution for patient data accessibility and decision support in healthcare service delivery.

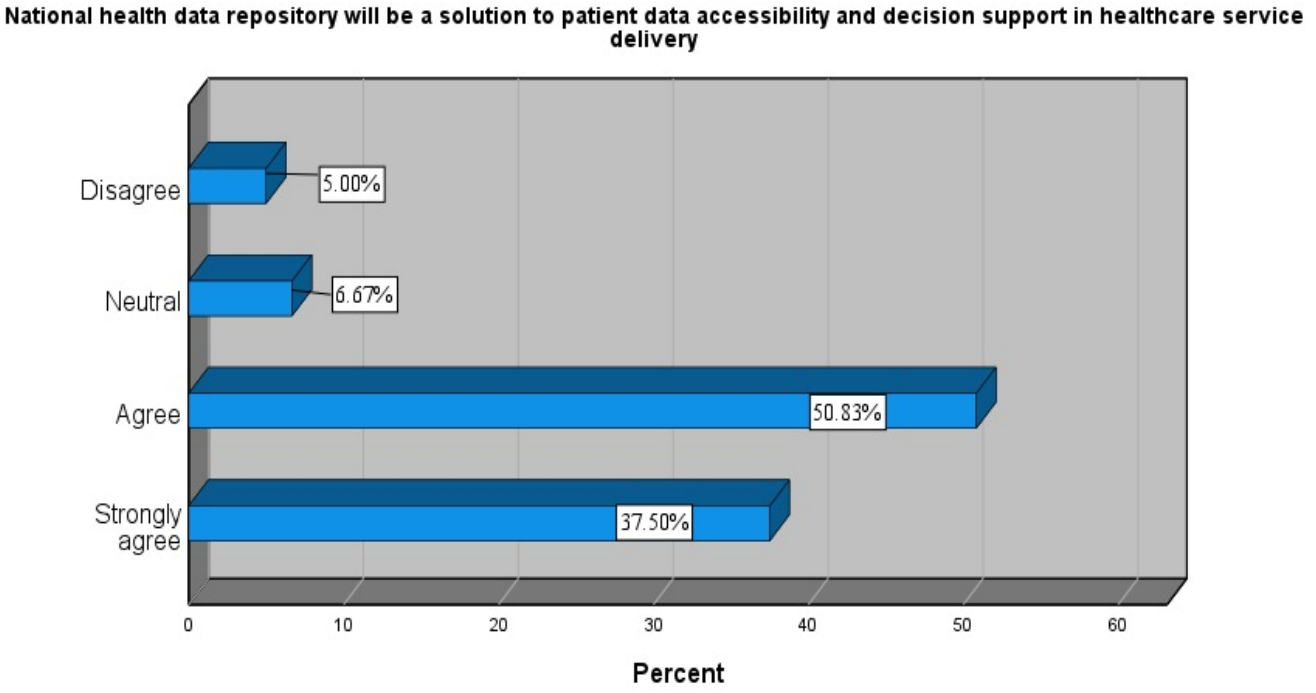

Figure 15. National health data repository as a solution to patient data accessibility and healthcare service delivery

The data analysis outcomes illustrate a prevalent positive sentiment among the participants. A substantial proportion, accounting for 37.5%, "Strongly agree" with the statement, indicating a robust backing for the notion that a national health data repository could effectively address the challenges of patient data accessibility and decision support in healthcare service delivery. Additionally, 50.8% of participants expressed "Agree," underlining most participants who are in favour of this concept. A smaller yet notable

subset of participants, at 6.7%, maintained a "Neutral" stance, representing a minority who withheld a strong opinion. In contrast, a mere 5% of participants responded with "Disagree," suggesting a small fraction of participants who hold reservations about the effectiveness of a national health data repository in resolving patient data accessibility and decision support challenges in healthcare service delivery. Importantly, no participants "Strongly disagreed," which underscores the overall lack of strong opposition to the proposed solution. These findings emphasize the generally positive perception among participants that a national health data repository holds promise as a means to address issues related to patient data access and decision-making support, reflecting a prevailing viewpoint that could potentially inform future healthcare data management strategies.

4.3.2 Clinical benefits of having a national health data repository.

The graph below visually illustrates our analysis of participants' responses concerning the perceived clinical benefits of having a national health data repository and Single Patient Record.

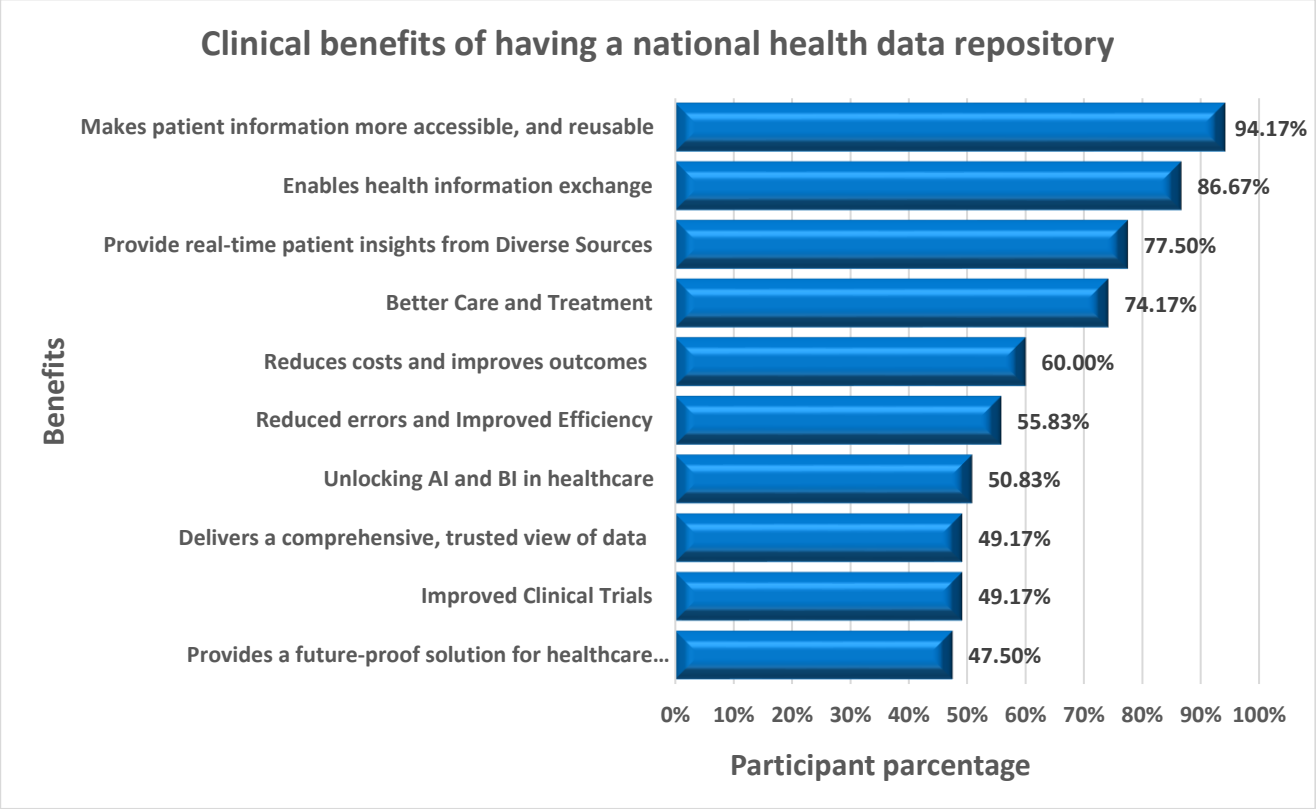

Figure 16. Clinical benefits of having a national health data repository

The graph provides an informative overview of participants' opinions regarding the potential clinical advantages associated with the implementation of a national health data repository and Single Patient Record.

The data analysis outcomes reveal a spectrum of opinions among the participants regarding the clinical benefits. Notably, "Makes patient information more accessible, and reusable" garnered the highest level of endorsement, with 94.17% of participants recognizing its potential. This underscores the paramount importance participants attribute to easy access and reusability of patient information. Subsequently, "Enables health information exchange" garnered a significant level of support, with 86.67% of participants affirming its value in facilitating seamless data exchange across healthcare entities. Other notable endorsements include "Better Care and Treatment" at 74.17%, "Provide real-time patient insights from Diverse Sources" at 77.50%, and "Reduces costs and improves outcomes" at 60.00%. Meanwhile, "Improved Clinical Trials," "Delivers a comprehensive, trusted view of data," "Unlocking AI and BI in healthcare," and "Provides a future-proof solution for healthcare interoperability" garnered slightly lower levels of endorsement, ranging from 47.50% to 50.83%. These findings provide a comprehensive understanding of participants' perspectives on the potential clinical advantages, shedding light on areas of consensus and differing opinions.

Overall, the outcomes highlight the perceived potential for a national health data repository and Single Patient Record to bring about substantial clinical benefits, suggesting a promising avenue for enhancing healthcare services through improved data management, accessibility, and reusability. Additionally, it's heartening to observe a substantial percentage acknowledging the repository's role in improving care quality. While the responses demonstrate varying degrees of recognition for other benefits such as health information exchange, real-time insights, and AI/BI integration, they collectively indicate a positive stance towards the repository's potential contributions to healthcare. The moderate recognition of certain aspects like improved clinical trials, comprehensive data views, and future-proof interoperability solutions might signal areas where further education and advocacy could bridge perceptions and elevate awareness among healthcare professionals.

4.5 Barriers against implementing national health data repository.

The graph below visually presents our analysis of participants' responses concerning the perceived barriers against the implementation of a national health data repository. The graph offers a

comprehensive overview of participants' opinions on the challenges that could potentially hinder the successful establishment of a national health data repository.

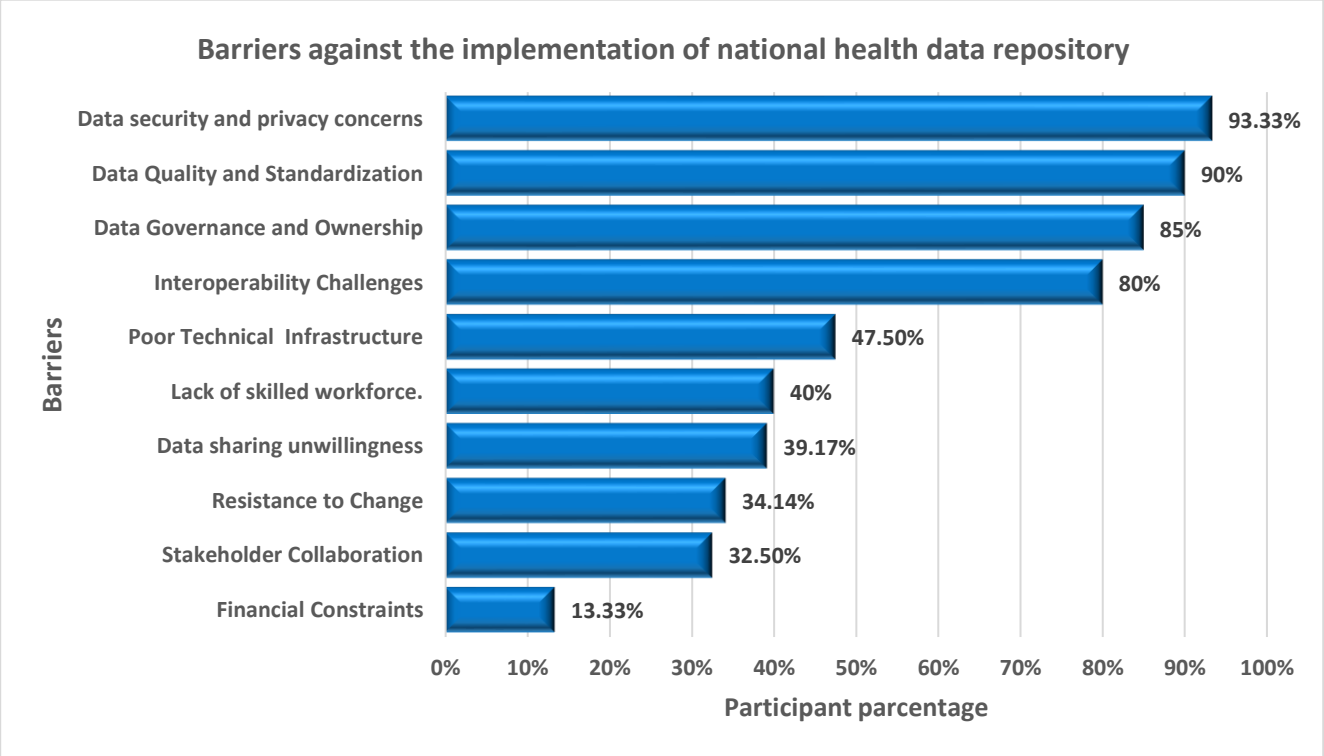

Figure 17. Barriers against the implementation of national health data repository

The data analysis results highlight a range of barriers participants recognize. Notably, “Data security and privacy concerns” garnered the highest level of endorsement, with a significant 93.33% of participants acknowledging its potential as a major obstacle. This underscores the paramount importance of safeguarding patient data in any data repository implementation. Similarly, "Data Quality and Standardization" received substantial endorsement at 90% while “Data Governance and Ownership” gathered substantial endorsement at 85%, reflecting the participants' recognition of the critical role these aspects play in a successful repository. Additionally, “Interoperability Challenges” was recognized as a significant barrier by 80% of participants, indicating the importance of seamless data exchange across various healthcare systems.

Other notable barriers include "Poor Technical Infrastructure" at 47.50%, "Lack of skilled Big data management and processing workforce" at 40%, and "Data sharing unwillingness among health institutions" at 39.17%. Meanwhile, barriers such as "Stakeholder Collaboration," "Resistance to

Change," and "Data Governance and Ownership" garnered relatively lower levels of endorsement, ranging from 32.50% to 34.14%. The relatively minimal recognition of "Financial Constraints" as a barrier (13.33%) suggests that participants perceive it as less inhibitive. These findings provide a comprehensive overview of the challenges perceived by participants, emphasizing the multifaceted nature of barriers that need to be addressed for the successful implementation of a national health data repository.

4.6 Best approaches for the successful implementation of a national healthcare data repository

The survey results present two primary approaches for establishing a central data repository linked to hospitals.

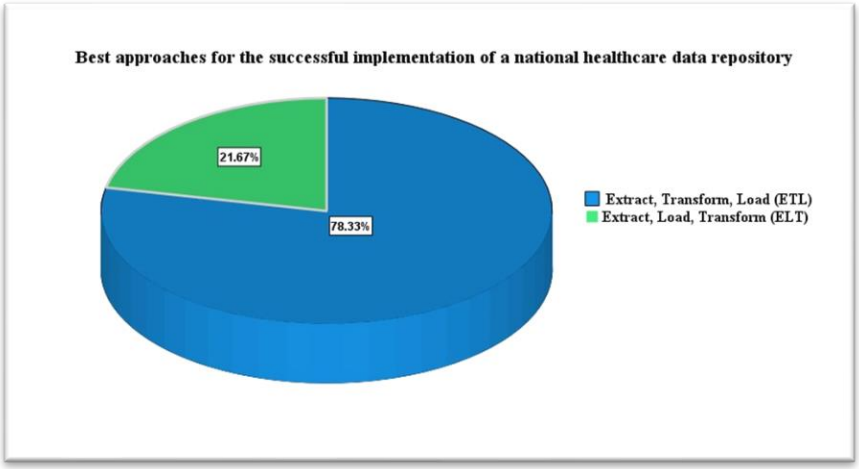

Figure 18. Best approaches for the effective implementation of a national healthcare data repository

The analysis of respondent preferences regarding the optimal approach for the successful implementation of a national healthcare data repository reveals a notable disparity. A significant majority, accounting for 78.3% of participants, endorsed the idea of a "Central data repository that is connected with all hospitals using Extract, Transform, Load (ETL) Technology." This approach, leveraging ETL technology, is favored by the majority of participants as a strategic means to ensure the successful implementation of a national healthcare data repository. This preference suggests a prevailing inclination towards a method that emphasizes data transformation and cleansing prior to loading, likely driven by the recognized benefits of enhanced data quality and consistency.

In contrast, a smaller yet notable proportion, comprising 21.7%, advocated for a "Central data repository that is connected with all hospitals using Extract, Load, Transform (ELT) Technology." While this approach garnered less support than the ETL counterpart, it still reflects a meaningful perspective among participants. These findings underscore the significance of advanced data integration methodologies in the implementation of a successful national healthcare data repository. The outcomes provide valuable insights into participants' inclinations, which could inform decision-making processes when considering the best approaches for effective data repository implementation in healthcare systems. The survey results underscore the significance of carefully considering the merits and implications of each technology approach in the context of healthcare data management, while also highlighting the need for tailored strategies to accommodate diverse perspectives and requirements within the healthcare ecosystem.

CHAPTER 5. RESULTS DISCUSSION

In this chapter, researcher delve into a comprehensive discussion of the key findings derived from the data analysis of our study. The results provide valuable insights into various aspects related to participants' perspectives on healthcare data management, access, and the potential implementation of a national health data repository.

5.1 Participants Characteristics

The thorough analysis of our participants' demographic information has illuminated several intriguing aspects of our study. The age distribution unveils a diverse sample, primarily clustered within the age range of 31 to 40 years, suggesting the presence of a considerable number of mid-career professionals. Adding depth to this understanding, a gender breakdown indicates that the majority of respondents were male (64.17%), while females accounted for 35.83% of the sample similarly to the study conducted in Uganda entitled achieving optimal data impact in rural African health care settings, the participants were 75% of male and 25% of male as gender distribution. This distribution of gender provides a valuable context for interpreting the subsequent findings (41).

Moreover, the participants' professional backgrounds significantly contribute to the multidimensionality of our study. The prominent representation from the healthcare sector (83.33%) reinforces the study's relevance and potential impact within this domain. Concurrently, the presence of IT and data management professionals (16.67%) lends a diverse technical perspective, enriching the research's scope and enhancing its interdisciplinary nature. This combined representation of healthcare and IT professionals enhances the robustness of our study's findings, aligning with real-world scenarios where effective data management and accessibility are crucial for both sectors.

When considering the participants' education and experience, it is noteworthy that participants with bachelor’s degrees (64.2%) constitute the largest group, followed by those with master’s degrees (28.3%) and Doctorates (7.5%). This educational distribution contributes to the comprehensive insights our study provides, spanning varying levels of academic expertise. Similarly, the range of participants' years of experience, with a notable concentration within 6 to 10 years (39.2%) followed by those with 1 to 5 years of experience make up 33.3% of the sample, provides a dynamic backdrop for interpreting their perspectives on data access, repository implementation, and potential barriers.

In summary, this multidimensional analysis of age, gender, professional distribution, education, and experience underscores the richness of our participant pool. These factors collectively shape the diverse viewpoints and insights that emerge from our study's findings, culminating in a holistic understanding of the complex landscape of healthcare data management and the potential for a national health data repository.

5.2 The Needs and Level of Support for a National Healthcare Data Repository

The findings indicate that a significant majority (85%) reported challenges in accessing such data. This highlights a pervasive issue in healthcare systems, where data silos across facilities hinder seamless information exchange. Additionally, participants' perceptions regarding the impact of these challenges on care service delivery were examined. The results reveal a range of sentiments, with a notable proportion strongly agreeing (32.5%) followed by 60.8% of participants agree, that accessing referred patients' data poses challenges to care delivery. This attesting the challenge posed by the lack of a comprehensive patient data overview from other health facilities which can potentially impact care service delivery. These findings underscore the importance of addressing data accessibility barriers to ensure efficient care coordination and patient management across healthcare entities. A core finding of implementation of data access and use procedures in clinical data warehouses with the aim of a systematic review of literature and publicly available policies was the lack of concrete information available in the scientific literature and the lacking public accessibility of clinical data which is a similar challenge lead to clinical data warehouse (15).

Moreover, our analysis explored participants' satisfaction levels with their current patient data access experience. The results showed a diverse range of sentiments, with a significant proportion expressing dissatisfaction (44.2%) with the existing process while a larger portion, accounting for 31.7%, felt "Neutral" about their experience. This suggests potential inefficiencies or limitations in the current data access mechanisms. Addressing these concerns is crucial to optimizing patient care and clinical decision-making, as well as enhancing the overall healthcare experience for both providers and patients. Furthermore, the integration of patients' historical data with the current clinical information system was a key area of investigation. Participants exhibited varied opinions, with a significant percentage agreeing (18.3%) that the current system provides a holistic view of patients' historical data. However, a considerable proportion disagreed (50%) with this assertion, suggesting room for improvement in the

integration and presentation of historical patient information. A smaller subset, accounting for 11.7%, "Strongly Disagreed," indicating a strong reservation about the system's ability to meet their expectations in this regard. This underscores the importance of a comprehensive patient data repository to enable informed decision-making and effective service delivery.

Participants overwhelmingly expressed support for the idea of developing a national health data repository, with 73.3% strongly supporting and 23.3% slightly supporting the notion. This resounding endorsement underscores the recognition of the potential benefits such a repository could offer, including improved data sharing, enhanced decision-making, and better patient care. The absence of participants opposing the idea further emphasizes the perceived importance and relevance of a national health data repository within the healthcare landscape.

Additionally, the correlation analysis conducted between participants' years of experience and their perceived need for a national health data repository and integrated patient records revealed a significant positive correlation. This suggests that professionals with more experience tend to recognize the value and necessity of comprehensive health data systems more strongly. This correlation underscores the potential influence of professional experience on perceptions of data management and integration, indicating a need for tailored strategies to engage healthcare professionals at different stages of their careers.

In conclusion, the analysis of participants' perspectives on data access challenges, satisfaction with current data systems, support for a national health data repository, and the correlation between experience and data repository need, collectively paints a comprehensive picture of the complexities within healthcare data management. These findings illuminate critical areas for improvement, highlight the urgency of addressing data accessibility barriers, and underscore the potential transformative impact of a national health data repository on healthcare service delivery. The insights garnered from this study serve as valuable guidance for policymakers, administrators, and healthcare professionals striving to enhance data-driven decision-making and patient care in the Rwandan healthcare system.

5.3 Clinical Benefits of Implementing a National Health Data Repository

In this section, researcher discuss participants' responses regarding their level of support for the idea that a national health data repository could be a solution to patient data accessibility and decision support in healthcare service delivery. The findings reveal a significant positive sentiment among participants towards this notion. The overwhelming support, with 37.5% strongly agreeing and 50.8% agreeing, underlines the general agreement that a national health data repository holds the potential to address challenges related to patient data accessibility and decision-making support. Moreover, the absence of strong disagreement signifies a lack of outright opposition to this solution. These findings highlight the participants' recognition of the repository's role in enhancing patient data accessibility and quality decision-making, suggesting a promising avenue for healthcare service improvement.

In addition, researcher examined participants' opinions regarding the perceived clinical benefits of implementing a national health data repository and Single Patient Record. The graphical representation of participants' responses in *Figure.19* offers a clear overview of their perspectives on various potential advantages. Notably, participants widely acknowledged the significance of “Makes patient information more accessible and reusable,” with 94.17% expressing its importance. Additionally, "Enables health information exchange" garnered substantial support at 86.67%, emphasizing the potential for improved communication and coordination among healthcare entities. Other benefits, such as “Better Care and Treatment,” "Provide real-time patient insights from Diverse Sources,” and "Reduces costs and improves outcomes," received notable recognition, reinforcing the belief that a national health data repository could contribute to enhanced patient care quality, informed decision-making, and resource efficiency. While some benefits like "Improved Clinical Trials," "Delivers a comprehensive, trusted view of data," "Unlocking AI and BI in healthcare," and "Provides a future-proof solution for healthcare interoperability" received slightly lower levels of endorsement, they still showcased a positive inclination towards these potential outcomes.

These outcomes collectively underscore the participants' perception of the clinical advantages associated with the implementation of a national health data repository and Single Patient Record. The responses reflect a positive outlook on the potential contributions of such a repository to various aspects of healthcare service delivery, decision-making, and efficiency. While certain benefits may require further advocacy and education to garner stronger support, the overall response pattern signals an openness to

embracing data-driven enhancements in healthcare practices. Similarly, various benefits were identified during the development of a National Health Data Warehouse for Data Mining in Bangladesh. This study highlighted several advantages, including standardizing data across the organization, enhancing the turnaround time for analysis and reporting, facilitating easy data sharing, alleviating the informational processing load from operational databases, improving data quality and consistency, providing historical intelligence, reducing the cost of accessing historical data, and restructuring data to optimize query performance and support decision-making. These findings align closely with the benefits identified in the present study (30).

In conclusion, the analysis of participants' perspectives on the clinical benefits of a national health data repository demonstrates a general consensus on its potential to positively impact patient care quality, decision-making, and resource management. The results reflect the participants' optimism about the repository's potential contributions to healthcare, highlighting key areas where its implementation could lead to significant improvements. These insights provide a foundation for informed decision-making and strategic planning aimed at harnessing the potential of a national health data repository to elevate healthcare services and patient outcomes.

5.4 Barriers Against Implementing National Health Data Repository

In this section, researcher explored participants' perceptions of barriers that could impede the successful implementation of a national health data repository. The graphical representation of participants' responses in Figure.14 provides a comprehensive overview of their opinions on the challenges associated with establishing such a repository. The outcomes of this analysis underscore the multifaceted nature of the obstacles that must be addressed to ensure a successful repository implementation.

The data analysis outcomes unveiled a range of barriers that participants recognized as potential impediments. Notably, "Data security and privacy concerns" emerged as the most significant barrier, with an overwhelming 93.33% of participants identifying it as a major obstacle. This strong endorsement emphasizes the paramount importance of ensuring the confidentiality and protection of patient data throughout the repository's implementation. Similarly, "Data Quality and Standardization" received substantial recognition at 90%, highlighting the participants' understanding of the pivotal role these aspects play in ensuring the accuracy and consistency of healthcare data.

Furthermore, "Data Governance and Ownership" gained substantial endorsement at 85%, reflecting participants' recognition of the essential need for clear regulations and responsibilities in managing healthcare data within the repository. The substantial recognition of "Interoperability Challenges" by 80% of participants underscores the importance of seamless data exchange across different healthcare systems, suggesting the participants' awareness of the complexities in achieving data compatibility.

Other barriers, such as "Poor Technical Infrastructure" (47.50%), "Lack of skilled Big data management and processing workforce" (40%), and "Data sharing unwillingness among health institutions" (39.17%), highlight participants' recognition of the technical and human resource challenges that may hinder the repository's establishment. Conversely, barriers like "Stakeholder Collaboration," "Resistance to Change," and "Data Governance and Ownership" garnered relatively lower levels of endorsement, suggesting a comparatively lesser perception of these factors as inhibitive.

Importantly, the relatively minimal recognition of "Financial Constraints" (13.33%) as a barrier indicates that participants perceive financial limitations as less significant. This could potentially indicate that participants acknowledge the long-term value and benefits of a national health data repository, despite the initial investment.

Overall, the analysis offers a comprehensive understanding of the barriers participants foresee in implementing a national health data repository. The outcomes highlight the multifaceted challenges that must be addressed, ranging from technical and human resource concerns to privacy and data governance considerations. The findings provide a valuable roadmap for policymakers, administrators, and stakeholders, informing strategic planning and decision-making to overcome these barriers and ensure the successful establishment of a repository that holds the potential to transform healthcare data management in Rwanda.

5.5 Best Approaches for Effective Implementation of a National Healthcare Data Repository

This section delves into the analysis of participants' preferences regarding the most effective approaches for the successful establishment of a national healthcare data repository. The graphical representation of these preferences offers a clear overview of participants' inclinations, which can inform decisions surrounding the implementation strategy for the repository.

The analysis of participant preferences yields two primary approaches for establishing a central data repository linked to hospitals. The survey results depict a significant majority, accounting for 78.3% of participants, endorsing the approach of a "Central data repository that is connected with all hospitals using Extract, Transform, Load (ETL) Technology." This resounding preference for ETL technology suggests a prevailing inclination among participants towards a method that emphasizes data transformation and cleansing prior to loading into the repository. This preference is likely rooted in the recognition of the benefits associated with enhanced data quality, consistency, and accuracy, which are pivotal in healthcare data management. On the other hand, a substantial yet comparatively smaller proportion, comprising 21.7% of participants, advocates for a "Central data repository that is connected with all hospitals using Extract, Load, Transform (ELT) Technology." This preference for ELT technology reflects an alternative perspective among participants, highlighting their recognition of different technological approaches in healthcare data integration.

The findings underscore the significance of advanced data integration methodologies in the successful implementation of a national healthcare data repository. It is evident that participants' inclinations are shaped by their understanding of the implications and benefits associated with each technology approach. The strong preference for ETL technology emphasizes participants' recognition of the importance of data transformation and standardization as fundamental pillars for accurate and meaningful data analysis within the repository. Importantly, these outcomes hold relevance beyond the specific technology choice, as they underscore the need for tailored strategies that accommodate the diverse perspectives and requirements present within the healthcare ecosystem. The participants' preferences indicate the importance of establishing a repository implementation approach that not only aligns with the technological landscape but also considers the details of healthcare data management, governance, and interoperability.

In conclusion, the analysis of the best approaches for effective implementation of a national healthcare data repository provides a robust insight into participants' technological inclinations and their implications for data quality and management. These findings offer valuable guidance for policymakers, administrators, and stakeholders as they navigate the complex landscape of healthcare data integration, ensuring a robust and effective repository that serves as a cornerstone for enhanced patient care and data-driven decision-making in Rwanda's healthcare system.

CHAPTER 6: CONCLUSION AND RECOMMENDATION

6.1 Conclusion

In this concluding chapter, researcher present a synthesis of the extensive analysis conducted throughout the study, drawing from the insights gleaned from the data analysis presented in Chapter Two. The diverse perspectives and opinions shared by the participants have provided a multifaceted understanding of the critical aspects surrounding the implementation of a national health data repository in Rwanda's healthcare landscape. Building upon the key findings, this chapter outlines our conclusions and offers comprehensive recommendations that involve all stakeholders in order to pave the way for an effective and impactful implementation of a national health data repository.

The analysis of participants' responses has provided valuable insights into various dimensions of healthcare data management and the potential benefits of a national health data repository. The study's findings reveal a diverse range of viewpoints, reflecting the complex interplay between technological, operational, and organizational considerations within the healthcare domain.

The study highlights the significant challenges healthcare professionals face when accessing data of patients referred from other health facilities, with an overwhelming 85% of participants acknowledging such difficulties. The participants' acknowledgment of the challenges emphasizes the need for comprehensive solutions to ensure smooth data sharing and accessibility across healthcare entities. Similarly, the analysis of participants' satisfaction levels with their current patient data access experience underscores the urgent need for enhancements to the existing data access processes to better support efficient patient care delivery.

Furthermore, the positive sentiments expressed by participants regarding the clinical benefits of a national health data repository and Single Patient Record reaffirm the potential value of such a system in improving patient care, data accessibility, and decision-making. The recognition of barriers against the implementation of a national health data repository underscores the complexities associated with data security, privacy, quality, and interoperability. This necessitates collaborative efforts to address these barriers systematically. The participants' preferences for approaches to successfully implementing a national healthcare data repository underscore the significance of ETL focussing on data transformation and standardization methodologies, suggesting an emphasis on data quality and

consistency in healthcare data management. Therefore, the following conclusions was drown from the study:

Firstly, the study indicated the critical need for improved mechanisms and solutions to facilitate the seamless access and transfer of patient data between health facilities. The existing challenges in accessing patient data can lead to delays in treatment, miscommunication between healthcare providers, redundant tests, and compromised decision-making. By implementing more efficient and secure data sharing methods, healthcare professionals can have timely access to a patient's medical history and relevant information, leading to better-informed and more personalized care provision.

Secondly, study indicated that they should be an Efforts to enhance data interoperability and data-sharing standards between healthcare systems and facilities can significantly improve patient outcomes. This may involve adopting electronic health record (EHR) systems that can securely exchange data, adhering to standardized data formats, and ensuring robust data privacy and security measures are in place through the implementation of a National health Data Repository.

Thirdly, implementation of a national health data repository could revolutionize healthcare delivery by enabling seamless data exchange and collaboration among healthcare providers. Having access to a patient's complete medical history, diagnostic records, and treatment plans in a centralized database can significantly enhance care coordination and informed decision-making, ultimately leading to improved patient outcomes.

Lastly, the survey results depict a significant majority, of participants, endorsing the approach of a "Central data repository that is connected with all hospitals using Extract, Transform, Load (ETL) Technology." This resounding preference for ETL technology emphasizes data transformation and cleansing prior to loading into the repository. This preference is likely rooted in the recognition of the benefits associated with enhanced data quality, consistency, and accuracy, which are pivotal in healthcare service delivery. The strong preference for ETL technology is due to its importance in data transformation and standardization as fundamental pillars for accurate and meaningful data analysis within the repository.

However, it is crucial to address privacy and security concerns when establishing such a repository to safeguard sensitive medical information. Striking the right balance between data accessibility and data

protection should be paramount to ensure public trust and successful implementation. Repository could be a game-changer in healthcare, revolutionizing how patient information is managed and utilized to provide more efficient and effective medical care nationwide.

6.2 Recommendation

Based on the comprehensive analysis conducted, researcher offer the following recommendations involving all stakeholders to guide the successful implementation of a national health data repository:

1. **Policy and Regulation Enhancement**: Minister of Health (MoH), Rwanda Biomedical Center (RBC) and other institution involved to collaboratively develop and implement robust policies and regulations that address data security, privacy, and ownership concerns. Engage relevant government agencies, legal experts, and healthcare institutions to establish a clear framework that safeguards patient data and ensures compliance with international data protection standards. These must be done to address the barriers like "Data security and privacy concerns" emerged as the most significant barrier, with an overwhelming 93.33% of participants identifying it as a major obstacle. Similarly, "Data Quality and Standardization" received substantial recognition at 90%, highlighting the participants' understanding of the pivotal role these aspects play in ensuring the accuracy and consistency of healthcare data. Furthermore, "Data Governance and Ownership" gained substantial endorsement at 85%
2. **Interoperability and standardized data formats:** Foster collaboration among healthcare facilities, IT professionals, and data experts to establish standardized data formats and protocols that facilitate seamless data exchange. This would enhance interoperability, making it easier for different healthcare systems to communicate and share patient data effectively. This will facilitate seamless data exchange between different healthcare providers, making it easier for medical professionals to access comprehensive patient information and make informed decisions. Collaborating with technology experts and adopting best practices in data management will be instrumental in achieving these goals. To overcome the substantial recognition of "Interoperability Challenges" by 80% of participants underscores the importance of seamless data exchange across different healthcare systems, suggesting the participants' awareness of the complexities in achieving data compatibility.

3. **Public and Patient Engagement:** Involve patients and the general public in discussions regarding the implementation of the repository. Transparently communicate the benefits, risks, and mechanisms for data protection to ensure public trust and understanding.
4. **Stakeholder Collaboration:** Establish cross-functional working groups comprising healthcare professionals, IT experts, administrators, policymakers, and data scientists. Encourage ongoing collaboration to collectively address challenges, share expertise, and optimize the implementation strategy.
5. **Continuous Improvement:** Develop mechanisms for continuous feedback and improvement. Regularly assess the repository's impact on patient care, data accessibility, and decision-making, and iterate on strategies to address evolving challenges as identified in this study.

By incorporating these recommendations, all stakeholders can play an active role in the successful implementation of a national health data repository in Rwanda. The repository has the potential to revolutionize healthcare data management, enhance patient care, and drive evidence-based decision-making, ultimately contributing to improved health outcomes for the Rwandan population. As researcher embark on this transformative journey, collaboration, transparency, and adaptability will be paramount to realizing the full potential of a national health data repository.

6.3 Area for Future Research

While this study has shed light on several crucial aspects of implementing a national health data repository, there are several avenues for future research that could further enrich our understanding and contribute to the successful implementation of such a repository:

1. **Cross-Country Comparisons:** Compare the implementation and impact of national health data repositories in different countries to identify best practices, lessons learned, and contextual factors that influence their success.
2. **Data Analytics and Insights:** Examine the potential of advanced data analytics, including predictive modeling and AI-driven insights, to leverage the repository's data for proactive patient care, disease prevention, and healthcare resource allocation.
3. **Impact Assessment:** Conduct a longitudinal study to assess the long-term impact of a national health data repository on healthcare outcomes, patient satisfaction, and operational efficiency. This could provide valuable insights into the sustained benefits and areas for continuous improvement.

4. **Ethical Considerations:** Delve deeper into the ethical considerations surrounding patient data sharing and access. Investigate patients' perceptions of data security, privacy, and their willingness to share their health information across various healthcare facilities.
5. **User Experience and Training:** Explore the user experience and training needs of healthcare professionals when interacting with the national health data repository. Investigate how training programs and user-centred design could enhance the repository's usability and effectiveness.

6.4 Limitations

Despite the valuable insights gained from this study, it's important to acknowledge its limitations, which can inform future research and implementation efforts.

Sample Size and Representativeness: The study's sample size may not fully represent the diversity of healthcare professionals and stakeholders in Rwanda. A larger and more diverse sample could provide a more comprehensive understanding of perspectives.

Cross-Sectional Nature: The study is cross-sectional design captures a snapshot of opinions at a specific point in time. A longitudinal approach could provide insights into changes in perspectives over time.

Contextual Factors: The findings are specific to the Rwandan healthcare context and may not be directly generalizable to other countries with different healthcare systems, policies, and technological infrastructures.

Data Quality: The accuracy and completeness of the responses depend on the participants' understanding of the questions and their willingness to provide accurate information.

By acknowledging these limitations and addressing them in future research and implementation efforts, researcher can continue to refine our understanding and approach toward establishing an effective and impactful national health data repository.

REFERENCES

1. Healthcare data integration: how to combine data from multiple sources | EMR, Big Data Solutions, Data Management, HIPAA, Compliance Certification, Medical Software, Product Development Strategy, System Architecture, CI/CD, Healthcare | Cprime Studios [Internet]. [cited 2022 Jan 23]. Available from: https://cprimestudios.com/blog/healthcare-data-integration-how-combine-data-multiple-sources
2. National Health Information Repository and Data Warehouse (NHIRD) in South Africa - Health Information Systems Program [Internet]. [cited 2022 Jan 23]. Available from: https://www.hisp.org/?p=2748
3. World Health Organization. Primary Health Care Systems (Primasys). World Heal Organ [Internet]. 2017;1–48. Available from: http://www.who.int/alliance-hpsr
4. MoH. Rwanda Health Sector Performance Report 2019-2020. Minist Heal Rwanda. 2020;1–95.
5. Date S, Number TO, Name COR, Submitted EU, Hakiba S. USAID Rwanda Integrated Health Systems Activity (RIHSA) Quarterly progress report FY 2021 QUARTER TWO. 2021;
6. Amoroso CL, Akimana B, Wise B, Fraser HSF. Using electronic medical records for HIV care in rural Rwanda. Stud Health Technol Inform. 2010;160(PART 1):337–41.
7. Roodenbeke Ed, Lucas S RA. Outreach Services as a Strategy to Increase Access to Health Workers in Remote and Rural Areas: Increasing Access to Health Workers in Rural and Remote Areas. 2019;20894.
8. Health MOF. Standard Operating Procedures for Management of Routine Health Information at Referral / Provincial and District Hospitals (Public and Privates). 2019;
9. OECD. OECD Health Policy Studies Health Data Governance Privacy, Monitoring and Research: Privacy, Monitoring and Research [Internet]. 2015. 198 p. Available from: https://books.google.com/books?id=sw6nCgAAQBAJ&pgis=1
10. Harrison JH, Aller RD. Regional and National Health Care Data Repositories. Clin Lab Med. 2008 Mar 1;28(1):101–17.
11. Palvia P, Palvia S. The feasibility study in information systems: an analysis of criteria and contents. Inf Manag. 1988 Jan 1;14(5):211–24.
12. Shin P, Jones E, Jacobs F, Tuckson R. Building a national data repository to measure and improve health center quality. J Ambul Care Manage. 2010;33(4):307–13.

13. Kim Y, Lee B, Choe EK. Investigating data accessibility of personal health apps. J Am Med Informatics Assoc. 2019;26(5):412–9.

14. Krist AH, Peele E, Woolf SH, Rothemich SF, Loomis JF, Longo DR, et al. Designing a patient-centered personal health record to promote preventive care. BMC Med Inform Decis Mak. 2011;11(1).

15. Pavlenko E, Strech D, Langhof H. Implementation of data access and use procedures in clinical data warehouses. A systematic review of literature and publicly available policies. BMC Med Inform Decis Mak. 2020;20(1):1–13.

16. Nadkarni P. Clinical Data Repositories: Warehouses, Registries, and the Use of Standards. Clin Res Comput. 2016;173–85.

17. Paper W. Data Repository Guide.

18. Kohl S. European health data space proposal launched. Eur J Hosp Pharm [Internet]. 2022 Jul 1 [cited 2023 Aug 19];29(4):240–240. Available from: https://ejhp.bmj.com/content/29/4/240

19. Cimino JJ, Ayres EJ. The clinical research data repository of the US National Institutes of Health. Stud Health Technol Inform. 2010;160(PART 1):1299–303.

20. Eli Dino Santos AD, Enrique Tayag DA, Aragona J V, Esteban CD, Bernolia JT, Tan CL, et al. National Health Data Repository Framework As of 28. 2022;(March):1–82.

21. South African National Department of Health Health. National Digital Strategy for South Africa 2019-2024 B [Internet]. Department of Health, Republic of South Africa. 2019. 1–36 p. Available from: www.health.gov.za

22. Data Repository: What it is, Types and Guide | QuestionPro [Internet]. [cited 2022 Nov 5]. Available from: https://www.questionpro.com/blog/data-repository/

23. Data Repository: Overview and Advantages | Astera [Internet]. [cited 2022 Nov 5]. Available from: https://www.astera.com/type/blog/data-repository/

24. Uddin MA, Stranieri A, Gondal I, Balasubramanian V. Rapid health data repository allocation using predictive machine learning. Health Informatics J. 2020;26(4):3009–36.

25. Langkafel P. Big data in medical science and healthcare management: Diagnosis, therapy, side effects. Big Data in Medical Science and Healthcare Management: Diagnosis, Therapy, Side Effects. 2015. 1–248 p.

26. Hall ES, Greenberg JM, Muglia LJ, Divekar P, Zahner J, Gholap J, et al. Implementation of a Regional

Perinatal Data Repository from Clinical and Billing Records. Matern Child Health J [Internet]. 2018;22(4):485–93. Available from: http://dx.doi.org/10.1007/s10995-017-2414-9

27. Data Repository | Egnyte [Internet]. 2021 [cited 2022 Nov 5]. Available from: https://www.egnyte.com/guides/governance/data-repository

28. Benefits of a central data repository – Teleo Technologies Inc. [Internet]. [cited 2022 Nov 5]. Available from: https://www.teleoinc.com/benefits-of-a-central-data-repository/

29. Einbinder JS, Scully K. Using a clinical data repository to estimate the frequency and costs of adverse drug events. Proc AMIA Symp. 2001;154–8.

30. Wicaksana A, Rachman T. 済無No Title No Title No Title. Angew Chemie Int Ed 6(11), 951–952 [Internet]. 2018;3(1):10–27. Available from: https://medium.com/@arifwicaksanaa/pengertian-use-case-a7e576e1b6bf

31. Medina-smith A, Medina-smith A. A Self-Audit of the NIST Public Data Repository Using the CoreTrustSeal Trustworthy Data Repositories Requirements.

32. Hong L, Luo M, Wang R, Lu P, Lu W, Lu L. Big Data in Health Care: Applications and Challenges. Data Inf Manag. 2018;2(3):175–97.

33. Ardini MA, Pan H, Qin Y, Cooley PC. Sample and data sharing: Observations from a central data repository. Clin Biochem [Internet]. 2014;47(4–5):252–7. Available from: http://dx.doi.org/10.1016/j.clinbiochem.2013.11.014

34. Abouelmehdi K, Beni-Hessane A, Khaloufi H. Big healthcare data: preserving security and privacy. J Big Data [Internet]. 2018;5(1):1–18. Available from: https://doi.org/10.1186/s40537-017-0110-7

35. Hong L, Luo M, Wang R, Lu P, Lu W, Lu L, et al. Big Data in Health Care : What Is So Different About Was ist so anders am Neuroenhancement ? Open Access. 2018;1(2):122–35.

36. (NBD-PWG) NBDPWG. NIST Special Publication 1500-6 : NIST Big Data Interoperability Framework - Reference Architecture. 2015;6(June):1–62.

37. Van Panhuis WG, Cross A, Burke DS. Project Tycho 2.0: A repository to improve the integration and reuse of data for global population health. J Am Med Informatics Assoc. 2018;25(12):1608–17.

38. Guldi J. Big data. The Routledge Companion to Historical Theory. 2021. 430–445 p.

39. Healthcare Data Integration Services Company - Data Interoperability [Internet]. [cited 2022 Jan 20].

Available from: https://www.harmonyhit.com/services/data-services/integration/

40. Bocquet F, Campone M, Cuggia M. The Challenges of Implementing Comprehensive Clinical Data Warehouses in Hospitals. Int J Environ Res Public Health. 2022;19(12):1–6.

41. Miiro C, Ndawula JC, Musudo E, Nabuuma OP, Mpaata CN, Nabukenya S, et al. Achieving optimal heath data impact in rural African healthcare settings: measures to barriers in Bukomansimbi District, Central Uganda. Int J Equity Health [Internet]. 2022;21(1):1–13. Available from: https://doi.org/10.1186/s12939-022-01814-1

ANNEX

PART I: INFORMATION SHEET

INTRODUCTION

My name is NSENGIYUMVA Emmanuel, a graduate student in the Master of Health Informatics at the University of Rwanda. I am conducting a research study entitled **"Feasibility study on the implementation of a national health data repository to improve patient data accessibility and healthcare service delivery in Rwanda"** This study will play a vital role in health data management, data integration, and knowledge discovery, support single patient record, health information exchange and improve data-driven decisions in Rwanda. I will determine the best approaches for the successful implementation of a national healthcare data repository to improve the care delivered In Rwanda. Therefore, I would like to invite you to assist with this project by agreeing to be involved in this study. You can take 10-20 minutes to answer the attached questionnaire. Any data you provide shall be anonymous and confidential. It will be for academic purposes only.

Your cooperation is highly appreciated.

Yours faithfully,

NSENIYUMVA Emmanuel

Note: For any questions, contact me at +250788200917 or email: nsemmanuel13@gmail.com. If you have questions about your rights in the study, contact Prof. Niyoyita Jean Paul Tel: +250781575240 Program Coordinator M.Sc. of HI – CMHS / University of Rwanda.

PART II: INFORMED CONSENT

I ....., agree to participate in the research project on **"Feasibility study on the implementation of a national health data repository to improve patient data accessibility and healthcare service delivery in Rwanda."**

I understand that my participation in this study is entirely voluntary and that if I wish to withdraw from the study, I may withdraw and discontinue participation at any time without penalty or giving reasons for doing so. If I withdraw from the study, I understand this will not have any effect on my relations with the researcher or my work. I understand that I may not receive any direct benefit from participation in this study, but my participation may help others in the future. I understand that the information I give will be kept confidential to the extent permitted by the law. I have read and understand the explanation provided to me and I voluntarily agree to participate in this study.

Name ……..

signature of participant……

PART III: RESEARCH QUESTIONNAIRE

A. Background Information of Respondents

A1. **What is your gender?**

- ○ Male
- ○ Female

A2. **What is your age group?**

- ○ Below 20
- ○ Between 20-30
- ○ Between 31-40
- ○ Between 41-50
- ○ Above 50

A3. **What is your job background?**

- ○ IT Background
- ○ Healthcare Background

A4. What's your highest level of education?

- ○ Bachelor's degree
- ○ Master's degree
- ○ Professional Degree
- ○ Doctorate

A5. What is your function/Role …….

A6. What is your Specialty …….

A7. How many years of experience do you have?

- ○ Below 1
- ○ Between 1-5
- ○ Between 6-10
- ○ Between 11-15
- ○ Above 15

B. Questions About Participants' Perceptions of The Development Of A National Healthcare Data Repository And Single Patient Record

B1. How satisfied are you concerning your ability to access patients' data?

- ☐ Very dissatisfied
- ☐ Slightly dissatisfied
- ☐ Neither satisfied nor dissatisfied (Neutral)
- ☐ Slightly satisfied.
- ☐ Very satisfied

Remarks

B2. How strongly do you agree with the following statement: "**The current clinical information system allows me to have a holistic view of patients' data and help me in decision making to ensure quality service delivery"?**

- ☐ Strongly disagree.
- ☐ Slightly disagree.
- ☐ Neither agree nor disagree
- ☐ Slightly agree.
- ☐ Strongly agree.

Remarks:

B3. Have you ever faced any challenges accessing patients' data referred from another health facility?

- ☐ Yes
- ☐ No

B4. How strongly do you agree with the following statement: **"Accessing patients' data referred from another health facility is a challenge to care service delivery since we don't have a holistic view of patients' data"?**

- ☐ Strongly disagree.
- ☐ Slightly disagree.
- ☐ Neither agree nor disagree
- ☐ Slightly agree.
- ☐ Strongly agree.

Remarks:

B5. How would you support the idea of developing a national health data repository and single patient 'records?

- ☐ Strongly I don't support the idea.
- ☐ Slightly I don't support the idea.
- ☐ Neither support nor don't support the idea
- ☐ Slightly I support the idea.
- ☐ Strongly I support the idea.

Remarks:

B6. How strongly do you agree with the following statement: **"National health data repository will be helpful in healthcare service delivery"?**

- ☐ Strongly disagree.
- ☐ Slightly disagree.
- ☐ Neither agree nor disagree
- ☐ Slightly agree.
- ☐ Strongly agree.

Remarks:

C. Questions About Analysing the Perceived Clinical Benefits and Barriers against Implementing A National Health Data Repository.

C1. Do you think a national health data repository that supports single patient records will add value to your professional in-service delivery?

- ☐ Yes
- ☐ No

C.2 Using a scale of 1= not at all important to 5= very Important, please rate how the national health data repository and the single patient record would be important to you.

1	2	3	4	5
☐	☐	☐	☐	☐

Remarks:

C.3 How strongly do you agree with the following statement: **"National health data repository and the single patient record would be a solution to patient data accessibility and decision support in healthcare service delivery."**

- ☐ Strongly disagree.
- ☐ Slightly disagree.
- ☐ Neither agree nor disagree
- ☐ Slightly agree.
- ☐ Strongly agree.

Remarks:

C.4 What do you think will be the clinical benefits of having a national health data repository and Single Patient Record? Select all possible benefit.

- ☐ Better Care and Treatment
- ☐ Reduces costs and improves outcomes by consolidating patient' records
- ☐ Makes patient information interoperable, accessible, and reusable.
- ☐ Reduced errors enable clinical decision support and Improved Efficiency
- ☐ Enables health information exchange.
- ☐ Improved Clinical Trials
- ☐ Provide real-time patient insights from Diverse Sources
- ☐ Delivers a comprehensive, trusted view of data across the care setting.
- ☐ Provides a future-proof solution for healthcare interoperability.
- ☐ Unlocking AI and BI in healthcare
- ☐ Others, write other benefits in bellow box if not listed.

C.5 What do you think will be barriers against the implementation of national health data repository?

- ☐ Data sharing unwillingness among health institutions
- ☐ Poor Technical Infrastructure
- ☐ Lack of skilled big data management and processing workforce.
- ☐ Data security and privacy concerns
- ☐ Interoperability Challenges
- ☐ Data Quality and Standardization
- ☐ Stakeholder Collaboration
- ☐ Resistance to Change
- ☐ Data Governance and Ownership
- ☐ Financial Constraints

D. Questions About the Best Approaches For The Successful Implementation Of A National Healthcare Data Repository.

What do you suggest as the required approaches for the successful implementation of a national healthcare data repository?

☐ Central data repository that is connected to all health care facilities that use Extract, Transform, Load (ETL) Technology

☐ Central data repository that is connected to all health care facilities that use Extract, Load, Transform (ELT) Technology

☐ Others

ETHICAL CLEARANCE

UNIVERSITY of RWANDA

COLLEGE OF MEDICINE AND HEALTH SCIENCES

DIRECTORATE OF RESEARCH & INNOVATION

CMHS INSTITUTIONAL REVIEW BOARD (IRB)

Kigali, 14/11/2022
Ref: CMHS/IRB/510/2022

Emmanuel NSENGIYUMVA
Health Informatics Program,
Center of Excellence in Biomedical Engineering and E-Health (CEBE)
CMHS, UR

Dear Emmanuel NSENGIYUMVA,

RE: ETHICAL CLEARANCE

Reference is made to your application for ethical clearance for the study entitled ***"Feasibility study of Implementing a national health data repository in Rwanda to improve patient data accessibility and healthcare service delivery in Rwanda".***

Having reviewed your application and been satisfied with your protocol, your study is hereby granted ethical clearance. The ethical clearance is valid for one year starting from the date it is issued and shall be renewed on request. You will be required to submit the progress report and any major changes made in the proposal during the implementation stage. In addition, at the end, the IRB shall need to be given the final report of your study.

We wish you success in this important study.

Assoc. Prof. Stefan Jansen (PhD)
Acting Chairperson Institutional Review Board,
College of Medicine and Health Sciences, UR

Cc:

- Principal College of Medicine and Health Sciences, UR
- University Director of Research and Postgraduate studies, UR

APPROVALS FOR DATA COLLECTION

NDERA NEUROPSYCHIATRIC TEACHING HOSPITAL
BROTHERS OF CHARITY
P.O. Box: 423 Kigali - Rwanda; Tel: +250 722 827 364 / +250 788 827 364, Toll free 2575
Website: www.carnesnderahospital.rw E - mail :nderahospital@moh.gov.rw

Ndera, the 2nd June 2023
No:393/NNPTH/DG/2023

To NSENGIYUMVA Emmanuel
Student in Masters of Health Informatics/CEBE_UR
Phone: 0788200917
University of Rwanda

Re: / Approval permission for DATA Collection for academic purpose

Dear,

Reference is made to your letter dated 25th May 2023 that request of the permission for data collection on **"Feasibility Study of Implementation a National Health Data Repository in Rwanda to improve patient data accessibility and health care service delivery, Case study NNPTH".**

We would like to inform you that your request has been granted.

Note that all your findings must be shared by a copy in the hospital management before presentation at School.

Wishing you success

Director General
Brother Charles NKUBILI

CC:
- **Head of IT Service**
- **Department of Administration& Finance**
- **Department of Education, Research, CPD and Quality improvement**

REPUBLIC OF RWANDA

Gatonde on..01.../06...../2023

Ref No.148..GDH/2023/SEC

NORTHERN PROVINCE
GAKENKE DISTRICT

GATONDE DISTRICT HOSPITAL

Phone: (+250) 788276239-E-mail: gatonde.hospital@moh.gov.rw / Website: www.gatondehospital.gov.rw

NSENGIYUMVA Emmanuel
Tel: 0788200917

Dear Emmanuel,
Following your letter dated May 25, 2023 requesting for permission to conduct Research at Gatonde District Hospital;

I am pleased to inform you that your request has been granted. During your research you will work under guidance of Education &Research Committee in Gatonde District Hospital that is chaired by Dr. HARINDINTWARI Ildephonse (Tel: 0788708137).

Your also requested to comply with hospital policies and procedures including obtaining informed consent from participants, protecting their privacy and maintaining the security of collected data. Kindly choose your available date to start your research.

I wish you good luck in your research

MINISTRY OF HEALTH - GAKENKE DISTRICT - GATONDE HOSPITAL

Dr. DUKUNDANE Dieudonné
Director General of Gatonde District Hospital

REPUBLIC OF RWANDA

Nyamirambo, on 05 JUN 2023
No ..299/NDH.NYAM/2023

CITY OF RWANDA
NYARUGENGE DISTRICT
NYARUGENGE DISTRICT HOSPITAL

NSENGIYUMVA Emmanuel
Tel:0788200917
E-mail: nsemmanuel13@gmail.com

Re: Data Collection Approval

Dear Emmanuel,

Reference is made to your letter dated on May 25, 2023 requesting the authorization to conduct research on «*Feasibility Study of Implementing a National Health Data Repository in Rwanda to Improve Patient Data Accessibility and Healthcare Service Delivery* »

After examining your request by the Education and Research Committee of Nyarugenge District Hospital in its meeting held on June 02nd 2023, we have the pleasure to inform you that you are authorized to conduct your research in the Hospital from June 06th 2023.

In order to assure the accuracy of collected data you should share with the hospital your findings and recommendation before submission of the final report to your university

Sincerely,

IBITARO BY'AKARERE KA NYARUGENGE

Dr François Xavier UWIMANA
Chairperson of Education and Research committee