Addis Ababa University

School of Information Science

And

School of Public Health

Designing of a Web-Based Integrated Health Research Data Repository System for Ethiopian Public Health Institution

By Kidist Alemayehu

A Project Submitted to the School of Graduate Student of Addis Ababa University in the partial Fulfillment of the Requirement for the Degree of Master of Science in Health Informatics

January 2018

ADDIS ABABA UNIVERSITY SCHOOL OF

INFORMATION SCIENCES AND PUBLIC HEALTH

M.Sc in Health Informatics Program

Designing of a Web-Based Integrated Health Research Data Repository System for Ethiopian Public Health Institution

By Kidist Alemayehu

Name and signature of Members of the Examining Board

Advisers	**Signature**	**Date**
_____	_____	_____
_____	_____	_____
Examiner		
_____	_____	_____
_____	_____	_____

DECLARATION

I declare that the research project is my original work and has not been presented for a degree in any other university.

_____

Date

This research project has submitted for examination with my approval as university advisers.

_____

_____

Signature

Acknowledgment

It is a great pleasure for me to express my gratitude to Dr. Gashaw Kebede, who has been an excellent advisor. He has taught me a great deal about research project and guides me in a proper direction to achieve the objective of this research project. I am also grateful to my advisor Dr. Mesfin Addise for his bringing the research area to my attention, keen insight, and for his all rounded guidance and support.

I am equally thankful to Ethiopian Public Health Institute for assistance in granted this scholastic career and the Addis Ababa University for sponsoring this thesis Research. I greatly appreciate and acknowledge also the support received from my colleagues ICT team and research staffs of EPHI.

I would like to express my deeper appreciation for Meseret Ayano coordinator of Health Informatics Program, for her support, facilitation and encouragement throughout the study period. Finally, I am extremely thankful to my friends and family members for their help at various levels and ever growing kindness for me.

Above all, I praise the Almighty God, for the strength and endurance He gave me all the time.
Glory is to Him!!!

Contents

Acronyms

AIDS	Acquired Immune Deficiency Syndrome
CSV	Comma Separated Variable
DWH	Data Warehouse
DBMS	Database Management System
DAO	Data Access Object
DR	Data Repository
DSS	Decision Support Systems
EPHI	Ethiopia Public Health Institute
HIV	Human Immune deficiency Virus
HRDR	Health Research Data Repository
ICT	Information Communication Technology
ISO	International Organization for Standardization
IT	Information Technology
MS	Microsoft
NFR	Non Functional Requirements
OO	Object Oriented
OSS	Operation Support Systems
PHP	Hypertext Preprocessor
SOP	Standard Operation Procedure
SQL	Standard Query Language
TB	Tuberculosis
RDR	Research Data Repository
WHO	World Health Organization
UML	Unified Modeling Language
URL	Universal Resource Locator
XML	Extensible Markup Language
HTML	Hypertext Markup Language
XMI	XML metadata interchangeable
XIF	XMF Interchangeable facility

List of table

LIST OF FIGURES

Abstract

Background: Public health data is source health research, it is an asset and potentially useful for further research. Data collection is done for various health researches and stored, hence from time to time increasing amount data and becoming challenge to handle. This study presents the existing data storage utility problems like holding duplicated database, disorganized research resources, lack of accessibility and difficult of treacing. Since lack of powerful data management tools, there is no way of optimal storing to integrate the research finding with its research data. The aim of the designed system is to fill the gap of the existing data storage system and help to have an efficient data management.

Objective: The goal of this project is designing and developing health research data repository system that can improve the research data management of EPHI, including preserving and sharing research materials as an integral part of the research process support for the benefit of the researcher.

Methodology: The system designed in iterative and incremental development with object oriented approach. In order to collect requirements, data collection tools used (Face to face interview, document review and group discussion). Analysis and designing of the system was performed by unified modeling language tools and Microsoft Visio software. The system prototype is also developed to understand the designed system and system usability test was done to assess effectiveness, efficiency and satisfaction of the users on the developed prototype.

The new system can improve the existing research data and resources handling, data sharing and storage utility. It solves the challenges of data management based on the need of stakeholders.

CHAPTER ONE

INTRODUCTION

1.1. Background

An information system in an organization generally helps Operation Support System (OSS) and Decision Support Systems (DSS). OSS serves the need of running the daily operations of the business. DSS provide historical information for analyzing the business that important decisions can be made appropriately. Organizations have realized the importance of hidden information, which can significantly improve the quality of decisions. Moreover, unlike consumable resources, information as an organization's intangible asset can be reused over and over without losing its value (Orr, 2000).

Nowadays, the use of Information Communication Technology (ICT) is almost in all aspects of life. Similarly, healthcare domain has changed over time from a relatively stable to a dynamic one (Çeken, 2014). Health Information System (HIS) have evolved through different technologies applying informatics to health care creates the possibility of enabling "rapid learning" health applications that helps effectiveness in research processing and communication. Deploying ICT in health research environment has helped to improve use of data for health care planning (Haux, 2006). Like other developing countries, Ethiopia have to have these technologies as a crucial tool and means to eradicate poverty, speedup socio-economic developments, provide effective and efficient public service delivery.

Health Research Data Repository (HRDR) is a secure and confidential virtual research environment created to support health research projects; which is collaboration across research disciplines to health research data and metadata throughout their lifecycle to promote the secondary use and re-propose of research. Moreover to encourage and offer educational opportunities regarding research data management best practices. HRDR is to stimulate health related research while building a collaborative culture of respect relating to data management (Doiron, 2015).

For many years individuals in the research community have called for increased coordination and interoperability among data repositories to advance the use of informatics in health care. Due to the data be located in different databases, large amount of electronic medical data cannot fully utilized by researcher. Even when the organizations that collect and distribute biomedical data are willing to share, incompatible data formats or data interface create challenges for analyzing data across multiple datasets (Castro, 2009).

Research institutions have invested in research data repository and technical infrastructure to stable their data processes. Hence, data model design which describe data elements, the way these elements are stored and the relationship between elements in order to facilitate storage and retrieval of information and initial processes can be done once and maintained centrally (Hai Hua, 2011).The typical repository services are identifying, browsing, searching and downloading system functionality (Lee, 2017).

Basically, a database is carefully designed and constructed repository of facts, it is a part of a larger whole known as an information system which provides for data collection, storage, and retrieval. The information system also facilitates the transformation of data into information that allows the management of both data and information.(Coronel, 2009). Recently, there have been active exploration and development on the design and implementation of data management repository systems to meet the needs of data intense scientific research and discovery. Because of the evolving nature of this new trend related standards and practices are still being developed, there is a need to survey and understand current data management practice to address existing gap (Zhang, 2015).

Information transformation is one of the Federal Ministry of Health (FMoH) priority agenda and it emphasizes use of quality information for action and data interoperability between data sources. It also includes data sharing to promote effective data use for evidence based decision making (FMOH, 2015). The Ethiopian public Health Institute (EPHI) is a partner institutional organization under the FMoH mandated to conduct public health research in the country. It generates scientific information required in the development of better methods and techniques of enhancing disease management, prevention and control in the country (EPHI, 2016).

EPHI is working on research in public health including public health emergency management, establishing quality laboratory system, and training public health researcher and practitioners for best public health interventions. To improve the health of the general public of Ethiopian undertaking on priority health and nutrition issues for evidence based information utilization (EPHI, 2016). Hence the institute research data management system has to be utilizing research resources and usage in reuse, traceability and accessibility of data.

Developing systematic and secure health research data archiving in EPHI can ensure that valuable resources are available for answering public health questions and addressing the existing storage limitation. The one technical concept which is central to database management systems formalized description of the data contained in the database available to the programs that wish to use the data. Data Repository (DR) gives a complete and consistent data from multiple sources which can be easily understood and used in research applications when the data files grow very large, repetitive scans for data selection may become prohibitively slow during the data searching (Tossy, 2014).

The aim of designed system is to improve the current research data management system of EPHI. It is also helps to give users at all levels of the public health research integrated, secure and consistent data get from which they could report on and set their research needs more efficiently.

Object-oriented (OO) modeling technique was employed the designed system that becoming industry's leading standards in design, their effort on modeling would very likely make impact on data modeling as well. Unified Modeling Language (UML) designed as object-oriented DR modeling to specify, visualize and document models of software system (Naiburg, 2001). This system is web-based desktop and server applications means of greater access and more effective integration of health information communication from disparate computer applications and other information resource (Athina, 2010).

1.2. Statement of the problem

Advancing public health research requires data from a number of databases containing health information. A major problem hampering progress was the inability to optimize existing research data resulting from limitations in the access and sharing of data. Researchers were managing multiple disparate information which makes searching difficult in disconnected data source.

Research data were stored in a mixed environment (Microsoft SQL Servers, Excel, and Access) in different locations difficult to manage (Joseph, 2013). Health researchers have clearly identified data sharing as one of their top priorities and trying to get a sharing of data among different users where the data itself are crude and not always what the researcher wants but they need to be able to access the huge amounts of data (Broccolo, 2014).

EPHI is working on written data policies and procedures to govern data collection and access for research use. These policies are being ability to include data sharing agreements between the institute

and the researcher but the existing system is not support properly applied to use uniform way of data sharing for all users.

Large amounts of data are collected using various applications and stored for public health survey due to unorganized data in the storage, it is difficult for user to identify data sources. These data have exclusively been used to detect health problems in specific work populations but not reused for further research.

In the current system data storage devices hold heterogeneous data sources and commonly including relational DBMS, Excel, MS access, CSV and data are present in different locations on different platform under different schema (representation). There is no system of optimal storing data that helps to integrate the research finding with its research data which causes problem of data interoperability to reuse. Therefore, developing research data repository could solve the limitations of the existing system and to have an efficient data management information system is necessary.

1.3. Objective

1.3.1. General objective

The goal of this project is designing and developing health research data repository system that can improve the research data management of EPHI including preserving and sharing research materials as an integral part of the research process support for the benefit of the researcher.

1.3.2. The specific objectives

For the attainment of the general objective list of specific objectives are:

- To assess the existing system of EPHI
- To collect and analyze user requirements of the proposed research DR system
- To define functional and non-functional requirements of research DR system
- To design and develop a prototype of the system
- To evaluate the developed prototype

1.4. Scope

The process of the health research data repository is to stimulate both qualitative and quantitative health related research while building a collaborative culture of respect relating to data management and confidentiality (Challacombe, 2016). In this project only covers designing and developing a prototype health research DR for public health research at EPHI to enhance research resource utility.

Establishing public health research DR is very useful in the design of the platform that can help in solving the existing data administrative problems. Since the routine job of EPHI‘s data management is more of facilitating research data processing such that the project addressed the research and ICT department only.

1.5. Significance of the study

The system provides the different stakeholders various functionalities and simplifies accessing, sharing and interoperability of the public health research data. The major idea of DR is data readily and easily available.

- The primary motivation to implement a DR usually centers on improving data accessing mechanisms based on the regulation and policies of the institute for potential users. It helps to enable researchers to review and access the existing research results for fulfillment research gaps or a baseline of new study. And also it is possible to do comparison of disease trend and a retrospective analysis, provide more information on the previous events association with the latest information. Students’ (master’s, PhD, post-docs) from different health research program review related research work and access research data for their thesis

- Besides having data available for users from data administrative point of view, the DR will make it easier to control security and storage utilization. The system will provide longitudinal data and integrated research resources centrally that are currently available in different department and hand of individual researchers. This is users avail the secondary data information, avoid interoperability problem of data sources and acquire new data source, thus the system helps saving research process expensive and users could get remote access.

- The institute benefits from cost effective system method of keeping integrated public health data for further research, increasing research data handling quality and well organization of research resource that would motivate research interest of the researcher. Indirectly community benefits from the research finding dissemination timely and achieve good data processing tools in facilitating data sharing (what to share and what to preserve among data) that would be very important to repurposing.

- The system will simplify data processing and saving a research time acquire of researchers. The security issues affecting IT environment that enable committed, maintenance of the data should be controlled by the database administrator and responsible for ensuring the security and integrity of the database besides promoting educational opportunities regarding research data management best practices. It can be used as a baseline for data administrator's development of advanced repository system.

1.6. Organization of this paper

In the process of identifying the important contents to be developed and in designing the public health research data repository, this project is organized in different chapters. The first chapter deals with background of the study, statement of the problem, objective, scope and significance of the study.

The second chapter presents review of literature in the area of health research data repository and discusses related works that asses the practice of data repository technology. The third chapter describes a research methodology for this research project and discusses the techniques followed for data collection, requirement analysis and design of the proposed system.

The fourth chapters discuss about the result of requirement analysis of the designed system and assessment of the current system and identified issues are presented in accordance with the different information system tools. While the fifth chapter presents the developed platform prototype. Finally, chapter six presents conclusion and recommendation based on the research project system design.

CHAPTER TWO

LITERATURE REVIEW

The aim of the literature review is explore the available and existing type of information which had been covered by the various researchers. The literature was reviewed from journals, the internet, reference books, working, papers, reports and periodicals.

2.1. Overview and concept of health research DR

The scientific and socio-cultural demand for research in science, Research Data Repository (RDR) is expected to be relevance for researchers that support to share their research data. Its corresponding services are characterized by store a wide variety of file formats under different conditions for access and reuse. Compared to the storage of research data, activities concerning the standardization of repositories providing research publications are far more established. Disciplinary studies show a large diversity among RDR, thus shaping today's research data infrastructures. The practice in biomedicine shows in many of its requirements can also be seen other communities which of several hundred thousand scientists use these RDR (Pampel, 2013).

Academic and research institutions are now taking interest in the creation, dissemination as well as preservation of knowledge through the creation of IRs. Research institution has usual responsibility as producer of primary research to preserve and influence their constituents by means of sharing their academic activities. The trend has been modified as the emerging of a technology-driven society against a previous analogue nation. IRs is a digital medium created by various academic research institutions with the goal of uploading all the findings of research to the universal public for the further knowledge and utilization of information for the progress of a nation (Uzuegbu, 2012).

There have been several program advocates archiving and sharing of health research data. This increases the clearness and accountability of research and encourages its reliability in allowing researchers to extend analyses. Data collection is a significant and expensive part of research, databases can be used repeatedly increases the financial return on research asset by reducing data duplication. Despite these benefits, systematic data archiving and sharing are not yet the norm, especially in low and middle income countries. Many of health research databases are not well

managed or used even by primary researchers and often data are stored informally by institutions or individual researchers which is secondary use impossible. Systematic and secure data archiving can ensure that these valuable resources are available for answering public health questions (WHO, 2012).

The best global partnership for development is to synergize learning and research. Bridging the information gap has been a proposal adopted by all but not totally appeared. Some good researches have been done by Africans but are either out of rich to indigenes or are not widely available because of poor delivery and lack of access ICT medium to transport them. Therefore availability of ICT is a vital target in the forwarding research. Hence, workable and interoperable IRs is avenues of attaining MDGs in Africa (UN, 2003).

Fighting HIV/AIDS, malaria and other diseases are several in journals and other health publications, information sources for infected patients are scattered in various research publications done in many higher institutions in Africa. IRs has the potentials to synergize this information and make them publicly available (Uzuegbu, 2012).The current situation at many university medical centers regarding the management of biomedical research data leaves much to be desired. Suggestion of the German Research Foundation, regarding the qualified management of research data, put a possible solution that installing local a central repository for individual research data pools (Franke T, 2013).

The Faculty of Nursing at the University of Alberta, the HRDR is a secure and confidential virtual research environment created to support health related research projects, collaboration across research disciplines, health research data and metadata throughout their lifecycle and to promote the secondary use and repurposing of health research data. It is also promoting educational opportunities regarding research data management best practices and motivates health related research while building a collaborative culture of respect relating to data management and confidentiality (Challacombe, 2016).

The Canadian research councils formulated a harmonized Open Access that requires all peer-reviewed journal publications available online. Researchers are also required to deposit their research data into a relevant disciplinary repository after publication of research results, and they keep original datasets (Austin, 2015). Research data to be available for future use, long-term preservation would be goals at the time data are created the security measures researchers implement to protect data.

In the healthcare applications range from processing of very low level of data objects to very higher level of data object hence healthcare data must be identified in time with multiple degrees of accuracy. These characteristics raise inevitable special issues and fundamental differences in comparison with many other domain data. The nature of data and its use for clinical care and research bring specific demands (Rujirayanyong, 2006).

Countries are now trying to have data repository and sharing mechanism. If we take South Africa, there are several electronic record systems as seen related to the data warehouse debate. For the Tuberculosis electronic register and from some anti retroviral treatment electronic registries for AIDS patient, "extract, transform and load" functionality is established with the District Health Information Software (DHIS), the standard more technically advanced model is gradually being adopted (Kossi, 2010).

Africa need to create an environment that would support the development of repository. This includes placing emphasis on capacity building and advocacy for institutional repositories and adopting an approach that takes into account some of the major concerns (Justin Chisenga, 2006). So far there was no integrated health research data repository in Ethiopian health sector but currently the FMoH has given due attention in the health transformation plan to establish an integrated health data warehouse/repository for easy access and to make accessible to all stakeholders such that analytical reports will be produced by responsible agencies and disseminated. This designing health research data repository is believed to be one of the mechanisms to mobilize health information transformation in the country (FMOH, 2015).

2.2. Description of DR

Repository is a central place in which a data is kept and maintained in an organized way, usually in computer storage. In this system context, it is a place where public health research data are collected, it may be directly accessible to users or may be a place from which specific databases, files, or documents are obtained for further relocation or distribution in a network. A repository could be just the aggregation of data itself into some accessible place of storage or it may also means of ability to selectively extract data (Rouse, 2005).

Types of repositories are domain, discipline and institutional and their difference is the culture of the organizations that manage the repositories. As such a chemistry community may develop a domain repository; a crystallography community may operate a disciplinary repository; and a university may

run an institutional repository. IR may increase an institution's value by providing access to intellectual work produced by its communities. This access along with an emphasis on research data archiving and sharing from motivates many institutions to put their efforts into the development of IRs and services for research data (Lee, 2017).

Common repositories for non-public clinical data are less well established; however it is proposed that achievement of specific features in data repositories for the archival of non-public clinical datasets, to enable appropriate cross-linking of these datasets to articles quality standards (Hrynaszkiewicz, 2016). Health research data is data generated in the health data and defines discipline based on health research data repositories that maintain and archive data. This is researchers and faculty who seek information about potential repositories where they may be able to deposit their research data (Ucl Libraries, 2017).

2.3. Archiving models in DR

Data archiving and managing in repository centrally to which researchers shall to submit their databases with supporting documentation. Data are either dump to the repository or terms are agreed to govern limits on access. It is cost-effective infrastructure to train and keep highly skilled data archivists. Subject-focused repositories model involves archiving and access combined with the support of experts in several participating institutions. These diverse centers of expertise include the original research groups and the secondary use of databases that researcher's concerns loss of control and provide ongoing support. The portal model databases are archived in repositories specialize in specific research areas and support provided to users benefits from skill with the data being managed and work against interdisciplinary collaboration locate (Rani, 2012).

2.4. Common purposes of health research DR

Public health research data is generated from community based studies to manage this research data through organization disciplinary, DR system would help facilitates the research process simple and motivate the researcher deposit their data at common place. Some of the tasks:

- Store variety of file formats under different conditions
- Provide institutional visibility through access to collective public health research work
- Organize research information to effective resource management and remote access

- Enable reuse and repurposing of research data
- Archive and preserve research resources centrally
- Facilitate data sharing and initiate research collaboration
- Manage research output and share its valuable digital content (Health science, 2017)

2.5. Issues consider a DR designing model

Repository system should be a holistic process taking into consideration like, institutional policies and user characteristics; Existing technical infrastructure and skills current and future target. The assessment needs to be including within a broader context considering technical, organizational, and socio-cultural issues. Implementing repositories is significant that will enrich the existing information environment through the creation of new services and collaboration models (Peer, 2012).

Architecture of an organization's data repository reflects the performance measurement and business requirement of the organization not specific technology. Creating a repository features in today's technology that are scalable to accommodate expanding volume, contents and web services with file format migration, peer review and enables repurposing of content and services. Understanding the information chain and relationship among various forms of related information objects is becoming critical to create a usable and effective tool (Brody, 2007).

2.6. Development process of DR

Repositories modeling mainly integrated data storage and metadata. The goal of the repository is to provide a mechanism that allows organizations to use enterprise wide information for decision-making activities(Inmon, 2005).There are good reasons to demand that the unified data model used by a data repository to integrate health data be object-oriented. Research and practical applications over the last decades have sufficiently demonstrated to object-oriented representations are uniquely able to capture the attributes of designed artifacts in a natural and computationally efficient fashion. (Huang, 2013).

2.6.1. Database modeling

The common architectural modeling used in data repositories is database conversion, database synchronization and federated databases. Database conversion is similar to data integration in various source databases are fed through a database conversion engine with the aim of developing an integrated target database (Coronel, 2009). It is usually conducted on large volumes of data, where the

databases are mapped into universal model. Database managing refers to coordinating existing data occasionally; the volume of data involved is considerably less than in database conversion, also there appears to be problems in the integration of the data in the database. Federated databases are driven by specific queries. It is the most complex and requires much more in-depth work (Gavin, 2011).

2.6.2. Metadata modeling

Metadata modeling management, the technical requirement support information modeling which is data tracking and managing descriptions of the database in an integrated way. It's implementing a layer of control services on top of the DBMS, called a repository manager, and integrating with many tools. The result of this integration is a framework for metadata management, called a repository system (Omar, 2017).The basic format for the creation of metadata to be used in the development of data repository in essence of Dublin core metadata element set which is an open global standard and retrieval of digital resources (dig_metast.pdf, 2007).

Structural metadata indicates how different components of a set of associated data relate to one another. The most straight forward example is relational database metadata, it has own mechanisms for storing metadata. Tables of columns in each database, what tables they are used in, and the type of data stored in each column. In database terminology, this set of metadata is referred to as the catalogue (Green, 2009).

2.7. Repository structure

The major arrangements are, first the physical data database a central repository storing the data of the entire organization operational and metadata. Second the logical configuration is defined, third there must be designed the data, application, technical and support architectures needed for data storage implementation. Forth the data storage must be optimized according to the users' demands. To achieve this purpose, a carefully analysis has to be made concerning the requirements of these four architectures (Matei, 2007).

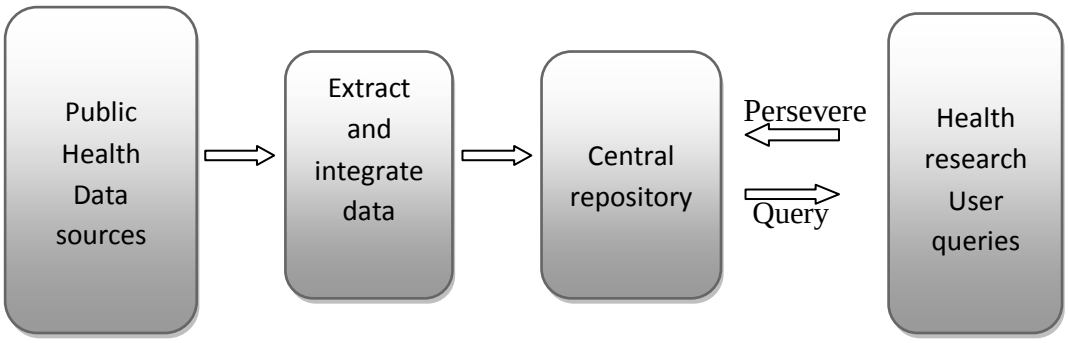


Figure 1: Conceptual Framework of the DR

2.8. The benefits of health research DR

Researcher can be able to put their research data with related documents in central place, which they do have memory of information of the data over time as people come and go, and their data meaningful to others (Creamer, 2014). This is one of the good opportunity researchers where can meet collaborators who may want to work with. The research data management systems automation becomes easy for research process and research data system integration of research workflows.

The data administrator perspective, able to see the data produced and released by different research unit in the previous years that can document productivity and better handling of reporting to decision-makers (Nurmela, 2015). Researchers feel using worth their time and continue using long-term that protected from irrelevant maintenance tasks and the data management plans they make to contain the whole lifecycle of the research data preserve the best way possible. Sustaining and trusted data repositories would increase the benefits from data collection task and methodical documentation available to other researchers. As well play a critical role in assuring that data remain accessible and available for future generations of researcher (Vardigan, 2015).

2.9. Limitation and challenges

Problems arise with data that are submitted without complete documentation and data files that lack a subject ID; incomplete archive of data, these omissions constitute a lack of origin for a sample which can create uncertainty about its quality and appropriateness for the reuse research effort (Pan, 2013).

In the process of developing and implementing the data archive, get encountered challenges in several areas: policy, technology, sustainability, extensibility and scalability and interoperability. These challenges require comprehensive responses, as cannot be solved by one academic unit alone. Need to be worked through within the broader context of the institution (Peer, 2012).

Sponsors and journals are increasingly requiring data sharing, but funding to ensure that data are adequately managed and preserved is often lacking. Some countries devote national funding for this purpose, but in others data repositories are left to fend for themselves often at the kindness of grant funding cycles, which do not assure sustained funding. There is concern in biomedical circles that some valuable databases have already been lost because funding ended communities do not have strategies for keeping data available (Vardigan, 2015).

2.10. Related works

International Network for the Demographic Evaluation of Populations and Their Health (INDEPTH) Data Repository is a unique resource for complete and detailed longitudinal public health data from geographically defined populations in low- and middle-income countries (LMICs) currently in 52 health and demographic surveillance system (HDSS) field sites situated in sub-Saharan Africa, Asia, and Oceania. INDEPTH is capable of producing reliable longitudinal data about the lives of people in the research communities. It is to enable INDEPTH member centers and associated researchers to contribute and share fully documented, high-quality datasets extraction, with the scientific community and health policy makers. The INDEPTH Data Repository to make data originating from the longitudinal surveillance conducted by its member HDSSs available to the scientific community. INDEPTH encourages its members to release a snapshot of this database on an annual basis as a core micro dataset on the repository. Datasets from multi-center studies conducted in INDEPTH member HDSSs are shared on the repository (Herbst, 2015).

Manitoba Population Research DR is a comprehensive collection of administrative, registry, survey and other data primarily comprised of residents of Manitoba. This repository is housed at the Manitoba Centre for Health Policy (MCHP). It was developed to describe and explain patterns of health care and profiles of health and illness, facilitating inter scrotal research in area of health care and other social services, the Manitoba Health Services Insurance Plan of all registered individuals. MCHP acts as a trustee or steward of the information in the Repository for agencies such as Manitoba Health (MH). Manitoba Population Research Data Repository reflects it's the expansion of research into areas beyond health care (University of Manitoba, 2016).

Government investment in Australia, eResearch infrastructure over the last several years provide all universities with an institutional repository for publications, and via the Australian National Data Service to encourage the creation of institution-wide Research Data Catalogues, and research Data Capture applications. The Hawkesbury Institute for the Environment, where a large sensor network gathers data for use by institute researchers, in-situ, with data flowing through to an institutional data repository and catalogue, and thence to Research Data Australia a national data search engine. A parallel workflow with a more generic focus any available to researcher and improve metadata capture at source, and create infrastructure that will support the entire research data lifecycle. This is advance tool for researchers to find, organize, package and publish datasets (Peter Bugeia, 2013).

NCBI is GenBank, an annotated online database of all publicly available DNA sequences. It acts as a central repository for genetic sequence data, exchanging data with multiple international partners on a daily basis, collecting sequence information directly from researchers, and receiving data. NCBI offers integrated search tools such as Entrez which searches NCBI's vast collection of biomedical databases and BLAST which allows researchers to find similar nucleotide or protein sequences in sequence databases. These tools link sequence information with related publications in databases. Data infrastructure could include linking current data sources to development of distributed electronic data networks and patient registries, and partnerships with the private sector (Castro, 2009).

2.11 Local related work

The AAU-ETD repository system is built on DSpace technology that is maintained and run by the Library System. Initially, it was a pilot project in specific disciplines later it was scaled up to include all subjects being covered by faculties of the University. Now it is a fully-fledged repository of about 2500 MSc and PhD theses and dissertations accessible to anyone in the world. Input to the system was at first handled centrally; then, following the decentralization of the theses submission system within the University, that allowed students to submit their work to the relevant departments, the Library had to change its approach accordingly. In management of the electronic theses collections, format control, metadata treatment and input into the repository. Users were provided with clear guidelines and secure remote access to put in the theses and can submit their own theses online, using the online submission interface (Wordofa, 2011).

2.12. Summary

Since EPHI is a research institute, it has usual responsibility as creator of primary research and preserve data for reuse but due to lack systematic storage utility and data sharing, the available research data are not yet potentially in use and some good researches results are not widely available. Hence learning from other countries research institute regarding the qualified management of research data, now they are trying to have repository and sharing mechanism. Designing web based integrated health research data repository system is relevant for EPHI to have efficient and secure data archiving that to make data originating from the longitudinal surveillance data available to further research and also good chance researchers where can meet collaborators who want to work with and easy to access their research data. It is data storage and metadata modeling management system with the technical requirement support information modeling which are traceable and managing descriptions of the research resources in an integrated way.

CHAPTER THREE

METHODOLOGY

3.1. Study area and period

This research project was conducted in Addis Ababa city administration, Ethiopia. The study was took place from January 2017 to December 2017 GC at EPHI. EPHI was restructured and has got its name Ethiopian Public Health Institute, by the council of ministers regulation No301/2013 which recognizes the institute an autonomous federal government office having its own legal personality and it is accountable for the FMoH.

There are eight hundred staffs in EPHI under different departments, HIV and TB, nutrition and food science research, technology transfer and research translation, health system, and including public health emergency management for best public health interventions. These are departments which performed evidence based research to improve the health of the general public of Ethiopian undertaking on priority health and nutrition issues (EPHI, 2016).

3.2. Study design

Designing of the system used an iterative and incremental methodology with object-oriented approach. This design methodology is chosen because of it emerges in response to the weakness of waterfall methodology (Ambler, 2005) that helps to improve the system step by step in a cyclic way to it satisfies user and approach to speed up the analysis, design and implementation stage (Dennis, 2012).

3.3. Source population and study units

Develop integrated research data repository in public health research data management of EPHI were considered in this project. Accordingly, source populations were all EPHI's staffs and the actual study units were those who are directly involved in research and data administrative departments with including management team of the institute.

3.4. Sampling procedures

Since this research project approaches product oriented, qualitative research method is convenient that would help addressing essential system requirement inputs. Purposive sampling selection technique was applied and data summarized manually by project researcher.

3.5. Requirements gathering

Semi-structured in-depth interview guide was developed to collect system requirement inputs. Face to face interview, on site observation and document review techniques were used for requirements gathering and analysis task. In addition focus group discussion was held to obtain a group opinion.

Interview is a two way method which permits an exchange of ideas and information (Pandey, 2015). It guides addressing various issue in the existing system like demographic data of respondents, about data use, description and storage, software's commonly used for research data process, users' commitment and responsibility of data sharing, and recommendation. The responses were marked on paper which was used as an essential input for the system requirement definition. Questions of the questionnaire are adopted from literature review of related work (hank, 2016).

Observation was one of requirement gathering tool, it is a technique of evaluation in which behavior are observed in a natural situations (Pandey, 2015). The existing system was physically observed system setting of the data management, data storage preservation and, system hardware and software that are currently in use. It also assessed how the existing environment helps to the designed system application.

The other tool used document review; it is collection, review, examination, and analysis of various forms of text as a primary source of research data (O'LEARY, 2004). Data access agreement, guideline and policies reference documents that were used for application of public health research and rules of the institute.

The interview was performed after asking the consent of the respondents whether they are willing to participate in this study and a consent request form was provided before they gave any information.

3.6. Requirement gathering and analysis quality assurance

Requirement gathering instruments were prepared according to the information need of the project before the requirement analysis was held. Requirement gathering compiled and checked for completeness was performed by researcher.

3.7. Object-oriented processing, design and analysis

The study applied an iterative and incremental with object-oriented analysis and design (OOAD) approach and recognized data interface and design standards. It helps development of quality monitoring DR technology task which integrated public health research and preserves historical data over time. System requirement analysis and design cycles were done using the UML tools of object oriented:

- Context diagram
- Use case diagram
- Class diagram
- Sequence diagram / state diagram
- System deployment

In the development of the prototype for this project HTML was used for front end construction of the interface and PHP was used in middle ware to create the different functionalities, business rules and connections to the database. MySQL server used to create tables and store various data. The tools were selected for the designed and development of the system majorly chosen on the basis of their ease to use, availability, supportability of the system environment and researcher prior experience. All of the tools are an open source which does not require any prior payments.

3.8. Evaluation technique

The new system evaluated based on the predefined objectives and user requirements in International Organization for Standardization (ISO) usability measuring aspects of system's user interface. The different parts of a system on a set of design principles were tested. The evaluation of interface designed system prototype was used heuristic evaluation technique and used self-administrative questionnaires to understand user preferences about the system designed prototype.

3.9. Ethical consideration

Ethical clearance was provided from Addis Ababa University School of Public Health to conduct this project. Official letter was provided to the EPHI and a consent form was given to all of the respondents prior to giving any information for the requirement collection.

3.10. Dissemination of results

At the end the project will be submitted to AAU as partial fulfillment of MSc in health informatics and communicated to Ethiopian Public health Institution.

3.11. Operational definition

Public health: the science of protecting and improving the health of families and communities through promotion of healthy lifestyles, research for disease and injury prevention and detection and control of infectious diseases.

Research: Scientific investigation that involves the generation of new knowledge or development of new theories, its result often cannot be applied directly to specific clinical situation.

Database: A computer application whose sole purpose is to store, retrieves, and modifies data in a highly structured way.

Data Repository: The term data repository means a database or a collection of databases that have been created or organized to facilitate the conduct of multiple research protocols, including future protocols not yet visualize.

Data: The term "data" means information derived directly from patients or human research subjects or indirectly through accessing databases.

Data warehouse: is an extract of an organization's data often drawn from multiple sources to facilitate analysis, reporting and strategic decision making.

Data dictionary: is a collection of descriptions of the data objects or items in a data model for benefits of programmers and others who need to refer to them. (The type of data such as text or image or binary value is described, possible predefined values are listed, and a brief textual description is provided).

Dataset: A group of related information comprising separate data elements stored retrieved or otherwise organized and treated as a unit, database itself can be dataset.

Extensible Markup Language: (XML): A cousin of HTML that provides structured data exchange between parties. XML contains data and metadata but no formatting information.

DBMS: Database management system based on the relational model that supports the full range of standard SQL. Uses a series of joined tables with rows and columns to organize and store data.

CHAPTER FOUR

ANALYSIS AND DESIGN OF THE SYSTEM

4.1. Analysis

4.1.1. Introduction

One of the essential steps of the system development is requirement collection and analysis. Assessing the existing system using various means is to provide necessary inputs for the system to be designed. The inputs should be organized in a way explore the functionality of the system that can express the flow of information, data handling application, and users of the system (Shelly, 2012). In this chapter the existing system in use, the business process, the system modeling and the design of the new system are presented.

4.1.2. The Existing System

The data repository technology aims to structure the data and database in appropriate way to access the data and use it in an efficient manner (Pampel, 2013). One of EPHI's business processes focuses on priority disease research and strengthening the national public health emergency services in the country. Research outputs from the institutes contribute towards the health sector development that are addressing major health problems of the communities and strengthen research capacities of national institutions and scientist in collaboration with responsible agencies (EPHI, 2016).

Currently the institute is often carrying out surveys, surveillances and intervention that apply to all fields of research for the public health benefit. Research has been done in different departments under disease research and public health emergency services core process of the institute. To conduct a research, the researcher or research project team should submit research proposal on public health related problems for ethical assurance, and develop standard operation procedure (SOP). Subsequently the database administrator would store the collected data randomly in the central storage and keep backup.

From this stored data there is informal data sharing and data request system, for instance one of the research departments of the institute is nutrition, under this department researcher request data from

nutrition server in order to do further research or review the prevalence and trends of malnutrition among children using the demography, underweight data element. Similarly the other research departments also researchers reuse data in their research subject area.

Researchers will promote a research environment that recognizes the value of data and seek to provide appropriate access of communication system and service to make their research data quality and reusable. They would identify good data management as an important aspect of their research process.

One of the application areas includes integral of data across the institution and health service providers, available online research result helps to policy maker's decision timely action on evidence based. Hence the institute data management needs an information based resource preservation system and capability to put the information into the hands of users.

4.2. Findings of the current system

Semi-structured in-depth interview and focus group discussion were used to collect data from the current system users. The following issues were noticed by the respondents on different information system context.

Table 1: Sample for the designed system, January to December 2017

Departments	Total number of Respondent		
	Gender		
	Male	Female	total
HIV/Tb	4	2	6
Nutrition and food science	3	2	5
Technology transfer and research translation	1		1
Health system	4	2	6
Administrators (research wing)	1	1	2
ICT professionals	3	2	5
Total	16	9	25

Through face to face interview tool the required information gathered from 25 individuals who are selected purposively the total of 800 source units of the study, as shown in the above table 1.

4.2.1. Users of the current system

Public health disease and nutrition research staffs and research collaborators are users of the current system, including students' (master's, PhD, post-docs) from different health research program. Data management of the research process performed by the institute ICT team with the researchers. All users are well computer skilled and have advance technology exposures. Based on the finding user raise a question to the institute research data management to be an efficient system and harmonizing research environment in different prospective.

4.2.2. Data storage and sharing process of the current system

The assessment has been done concerning various data that the institute uses and information generated. The EPHI's research database is storing in different servers and preserves data by taking back-up. There are huge amount of data acquire from the institute research and external research collaborators.

There is informal data sharing within the institute and with external partners which is time taking due to lack of automated system. Identified problems related to data storage processes and usability's are:

- Lack of information about available resources in the storage
- Disintegration of research result with its data source
- Lack of accessibility
- Disorganized database and difficult of traceability in the storage
- Problem of data storage utility and duplication of database
- Inability to generate report at a central level
- Lack of systematic data sharing to promote secondary data use

Based on the observation and the respondent's response about the existing system, the available network infrastructure helps to use different technologies for their research environment but the current data handling system functionality is not fully support their research process systematical to make it cost effective.

4.2.3. Documentation of the current system

Data documentation ensure that data will be understood and interpreted by any user and it will explain how the data was created, the context of the data, the structure of the data, and any manipulations that have been applied to the data (Pandey, 2015)

According to document review there is documents in data access, data sharing policy and agreement, ethics of the research context and SOP research projects. But these useful documentations are not regularly applied for every research which conducted in the institute. Inaddtion from finding of user interviews, lack of culture strict rule of data sharing agreement and know how about data sharing and management are the major barriers to sharing research data.

4.2.4. Software and hardware in the current system

Issues noticed from the group discussion participants' opinion, the researchers need to access a regular set of analytic software, SPSS, SAS, Stata, MSoffice, and other statistical software packages are assessed and need implemented respects with technology update. Data sources are created independently and need to have different schemas in order to avoid integration problems that can occur in data sources and applications software's.

Infrastructure of the institute is well equipped in hardware and software; there are data servers, application server and web server. It is networking system with broad band internet access and these resources can support implementation of the designed system.

4.3. The proposed system

Development of the designed system is result of user requirements with the roles and responsibilities of the institute. One of future plans of the institute was to establish well-organized research resources management system and to enable data sharing portal based on research data access rules and policy which is potentially useful for further study and data reuse. The system will be valuable to exchange information across the researcher and health institutions without geographical and time limitation. This central data repository provides tools of tracing and sharing data easily, and handling data storage utility.

There are several initiatives at a national and international level the establishment of integration of clinical information systems within clinical research and public health systems (J. Bledsoe, 2012). These initiatives aim to overcome technical issues of stability and interoperability between clinical information health systems undertaking an efficient reuse of personal health data for health research and data sharing are crucial the development of public health and clinical research (Eder, 2012).

4.3.1 Functional requirements

Functional requirements majorly deals with explaining on what has to be done by identifying the necessary task, activity and functionalities of the system should provide to users and tasks that must be accomplished (Dodero, 2008). Besides those functionalities will enhance secure health data management proper control of service utilization and standard data definitions across the system. The requirements are utilized for data model development which offers the justification behind the development of data repository (Little, 2003).

The system organizes research data and related information at a central data repository. Integration of research results with its research resources and enables to have data sharing portal through establish search engine in the system. According to the finding, the functional requirements of the designed system are as presented in tables 2 and 3 below:

Table 2: Functional requirements back end of the system

Store	Descriptions
Research results info	The system shall enable capturing data about research results / outputs
Research data sources	The system shall enable capturing the research data sources
Application SW	The system shall enable capturing information application SW used for research data processing
Data request	The system shall enable capturing information of requested data
Data delivery	The system shall enable capturing information of data delivered
Data dictionary	The system shall enable capturing information of data source elements
Ethical clearance	The system shall enable capturing information of research ethical clearance
User account	The system shall enable capturing information of user account

Table 3: Functional requirements front end of the system

Functions	Descriptions
Register	The system shall enable users to register and capture research data
Searching	The system shall enable users to search research result by author, year, name
Request data	The system shall enable users to request data
Delivery data	The system shall enable users to deliver data
Generate report	The system shall enable users to generate report
Upload data,	The system shall enable users to upload data to the system
Import data source	The system shall enable users to import data source
Querying /extract	The system shall enable users to extract data from data sources
View	The system shall enable users to view research documents

4.3.2. None functional requirement

Non-functional requirements define the overall quality element of the resulting system and including functional behavior and quality that the product must have. These quality factors comprises of appearance efficiency property (IIBA, 2009). Mainly non-functional requirements focus on the visible

quality aspects of the system capability in terms of integrity, portability, scalability, reliability, performance, usability and usefulness, security and maintainability (Scotte M, 2004).

4.3.2.1. Integrity

The system enables handle accuracy and validity of data source. The appropriateness of the source information satisfies user need and traceability capacity of relating user requirement. The system allowing data of table to import in different easy readable format and handling interoperability file format.

4.3.2.2. Availability

The system will accessible by all users every time and the degree to which the system should work for users that specifics availability, downtime, time to repair, accuracy, etc.(Tian J, 2006). The system enables user considering aspects of reliable maturity, fault tolerance dependable, recoverability. It is establish on every time running servers with power generators.

4.3.2.3. Portability

The system has a goal of producing a health research data repository in order to provide online access of research output and data request option to health research domain without geographical and technological limitations. Functions enable to be suitable various platforms as it can the system adopt software.

4.3.2.4. Privacy and Security

Privacy and security of health related information important issue when implementing a data repository. The system enables to protect unauthorized access whether accidental or deliberate to program and data and capable of guarding against unauthorized disclose. The security involving user right and governance which is systems and people communicate with the right rule that refers to the ability to provide confidentiality, data integrity, data availability and privilege hierarchy (Jana S, 2015)

4.3.2.5. Easy to use

All error messages produced by the system are meaningful; the system employ a single set of user interface provide a familiar and common look for the application. The interfaces are simple, less time taking to complete a job and suitable for user's needs and support different accessibility types of control for users. Since the system involve users with different technology tool familiarity from health

research background and usage capability, user interfaces would reflect health scheme in consistence appearance.

4.3.2.6. Performance

Performances while trying to get certain information all important search and view results are clearly available (Brown BB, 2016). System provide proper adequate response for accessing each of the activity and able to bring information as per the request of the users. In simple search function the system shows quick response and also it is capable scalability option of hardware and software for further system expansion.

4.3.3 Business rule identification

Business rules are at the base of every information system as they drive for managing and conducting all activities within an organization both operational systems and logical ones. Business rules implementation solutions are presented and analyzed according to some basic criteria that must be taken into account when choosing a business rules implementation strategy (Andreescu, 2012). Business rules influence the designed system considers.

BR1: Validate Information
Description: validates whether the input fields in every registers are filled correctly and important fields are not missed and input fields should be filled by appropriate input types.

BR2: Check out the completeness of documentation.
Description: Information about the research result with its data source should be integrating with necessary its research documentation available in the repository.

BR3: Version control
Description: Updating documents interns of their created time.

BR4: Repository check in/checkout
Description: Every time the developer performs an update (a check out), they get the latest version of the code base and a developer has made changes to the code, by checking in these changes (or committing them) the developer adds their changes to the repository.

BR5: Access control for security

Description: Formal application of the institute data policy and rules for information access and data sources protected from unauthorized access.

BR6: Repository backup and recovery

Description: Preparing replicate system to overtake and the existing system to continue working when system failure happened.

4.4. Modeling

Modeling is the process of gaining deeper understanding of a system through imitation and reflects properties of the system. Models of dynamics specify what a system does and how it reacts to stimulus from its environment, and how it evolves over time (Claudius p, 2014).

4.4.1. Process modeling

A process model is the conceptual representation of the business process in the organization that showing the overall model of the data, decision logic in business processes and activity flow in the organization. It is used to communicate information regarding a process and the interaction includes within / between organizations. Process modeling can also be used to show the comparison between the current system and the newly proposed system (Dennis, 2012).

Conceptualization of the various functionalities and behavioral representations presents multiple rough levels of clarity. It is from simple description of the workflow to simulation and execution. This enables the transparency of the business knowledge with a view to agree and bind all stakeholders in a representation that is shared within an organization and is reflected in its information.

An organization's business processes are best described by means of a business process model. A business process is entirely independent of who carries out the activities and the organizational structure of the organization (Cousins, 2002). One of the business processes of EPHI performs public health research activities to improve the public health and nutrition issues on evidence based information utilization. This research process including data collation and store, from these data there is informal data sharing with data sharing agreement and polices as shown in the figure 2 below.

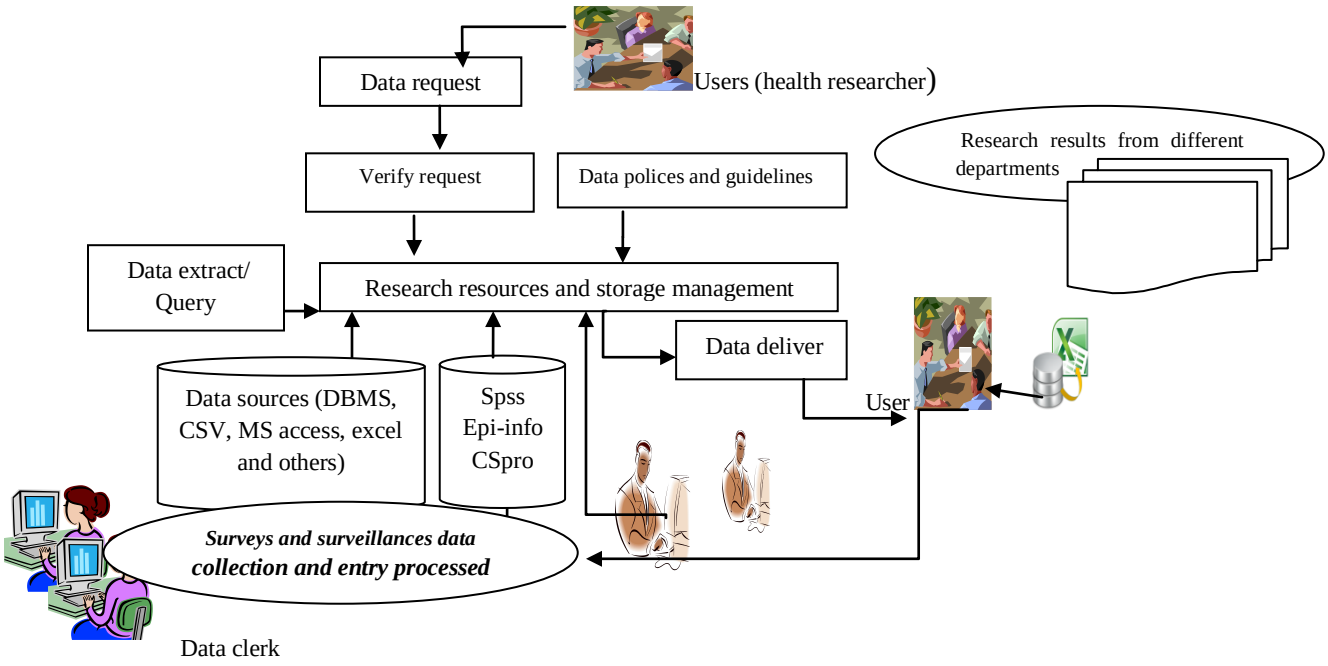


Figure 2: Business Process of Existing System Data Management

4.4.2 The public health research data repository system (To be)

The proposed system will integrate different functionalities that need to be operable for the full functioning as a system. As a repository is supposed to store raw and processed data as well as capture related information including services are identifying, browsing, searching and uploading system functionality. In order to hold up this functionality, the system bid a combination of automated metadata extraction and metadata entry possibilities. When it comes to research data reuse it cannot be performed directly in the system; data need to be extracting from the repository.

The importance of being able to search/query metadata is justified by the descriptive character of metadata. Researchers perform their queries from the system using search function and then the system display the result based on user request.

Incorporation of consistent interfaces of the system and functionalities like search/query and capturing information and display of metadata in coincidence with imported or entered data. Since, the stored

data with metadata would be easily managed for reuse, it would be very simple to find in the repository and its origin would likely be known, interpretation it effective for certain study processes. Researchers would like to use their own software, feel most comfortable with for the given task; therefore, more important than the actual integrated data viewer is the possibility to use other viewing software for the archived data.

The designed system has a home page that can be accessed by any anonymous user. In this page user can view research abstract, data policy, guidelines, and user can search research results. There is user registration and login form link. The system capture user registration data without duplication user's email address and avoiding null field.

Using system login form user can login to the user page and get access, which allows user to view research data sources information's (dataset, data dictionary, software, data source). User can request data online through data request submit in this system.

Using system login form administrative can login to the admin page which contain upload/download data tool, data storage form, view data request, data import, report generate, and user account manager system. In addition data delivery process performed in this page.

Many researches were done in public health at different time, where each research was collected data. This data are keep at different locations on different platform (schemas). Hence the system will facilitate to be stored as metadata with the associated data to all archived data.

4.4.3. Contextual modeling

This modeling is expressing the business process framework that fits into with all the external entities receive from and contribute to the information system. It assists to decide system boundaries; functionality included and provides a view of the overall business process as one process (Dennis, 2012). Contextual modeling of the study data repository system functionality hold user to registration in the system, storing information, uploading and data extracting with provide accessing research need (abstracts, guidelines and data polices) and data sharing environment. Inaddtion preserves integrated research result with its data sources and having searchable system as shown in the figure 3 below.

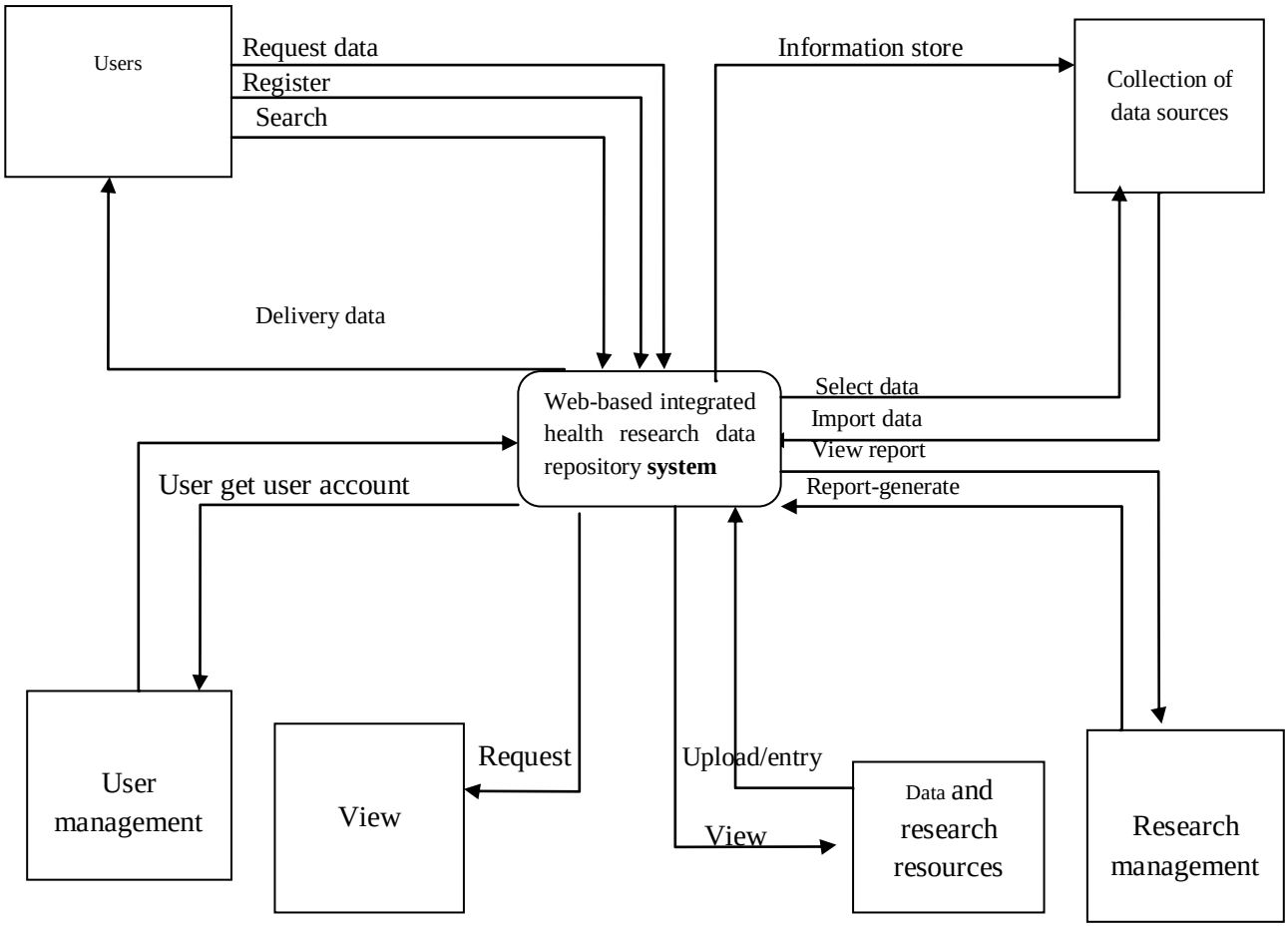


Figure3: Contextual Diagram for Proposed System

4.4.4 System modeling

Modeling is the processes of developing abstract of a system that allow represent a model in different perspective of a system. It is very important for readability and reuse approach of the systems. Various models are used during the requirements business process to provide interaction, structural and behavioral perspective of system. These are input during the design process to describe the system to implement and to prepare documentation of system structure and operation (Sommerville, 2004).

4.4.4.1. Object oriented UML modeling

There are good reasons to demand that the unified data model used in a data repository to integrate health data be object-oriented. Research and practical applications over the last decades have fully demonstrated that object-oriented representations are uniquely able to capture the attributes of designed artifacts in a natural and computationally efficient fashion. Object-oriented programming has by now established itself as the paradigm of choice for the development of robust, extensible and reusable software (Meyer, 1988). It representation is therefore not only desirable in its own right, but also facilitates data import to object-oriented applications as they are coming into wide use.

Object-oriented analysis looks at the problem domain, with the aim of producing a conceptual model of the information that exists in the area being analyzed. Analysis models do not consider any implementation constraints or how the system is to be built. The identified objects reflect entities and operations that are associated with the problem to be solved (Pamnani, 2015).

4.4.4.1.1 UML modeling

Unified Modeling Language (UML) is a very dominant graphical modeling language for specifying, constructing and documenting the artifact of software system. It provides a comprehensive notation for the full lifecycle of object-oriented development. (Pamnani, 2015). It becomes the standard modeling language for object-oriented modeling tool; the most common diagrams used being use case, class, sequence, state and activity (Omar, 2017). These are applied on the system modeling of central data repository.

4.4.4.1.2. Unified object modeling

An object-oriented (OO) representation is a collection of objects with attributes, the attributes that an object can have typically determined by the class to which the object belongs. Classes can inherit attributes from super classes and pass these attributes to all objects belonging to the class. An object model or schema is the collection of all classes describing the creation of communication for an application (Booch, 1999).

The unified object model underlying to in this study is a data repository scheme, its classes define the types of objects that can be instantiated to import data to health research clients. UML has become the standard language used for object-oriented software development (Naiburg, 2001). Notation used to develop and document the object model basic repository; before developed the schema, looking at the data how to manage in the central repository from various sources.

A DBMS, for instance, provides data in the form of flat tables with multiple columns recording various attributes of research objects. This is data models used in the data sources structured collections of value attributes and collection of classes. An infamous translation problem for object-oriented representations with possibly many-to-many object relations: the translator receives objects one-to-one, maps each to a representation understandable and sends it on. The simplicity of the resulting schema is not only adequate for the data it has to capture, but also greatly facilitates interpretation at the research data object is associated with one or more publication object (Koshak, 2002).

4.6.4.2 Use case modeling

Use case is the simplest and effective technique for modeling system requirements from a user's perspective. It contains information needed to build part of a process model in the system and can be identified by reviewing the functional requirements. An event response list also is useful in identifying the significant events that should be described in a use cased (Heumann, 2008). In the use case modeling identify functions of the system and target users (Actors). The tasks describes the actions of an actor when following a certain task while interacting with the system to be described in a name, number, importance level, brief description, primary actor, pre-conditions, post-conditions, major inputs and outputs, and scenarios of task. It is often new and extended functional requirements can be derive.

Table 4: Users and their role description in the system

Actor name	Description
User	Public health researchers /providers and decision makers
Anonymous user	User access only the home page or don't have user account
Database administrators	EPHI 's database administrators
Research and ethics committee	A team of research staff at EPHI

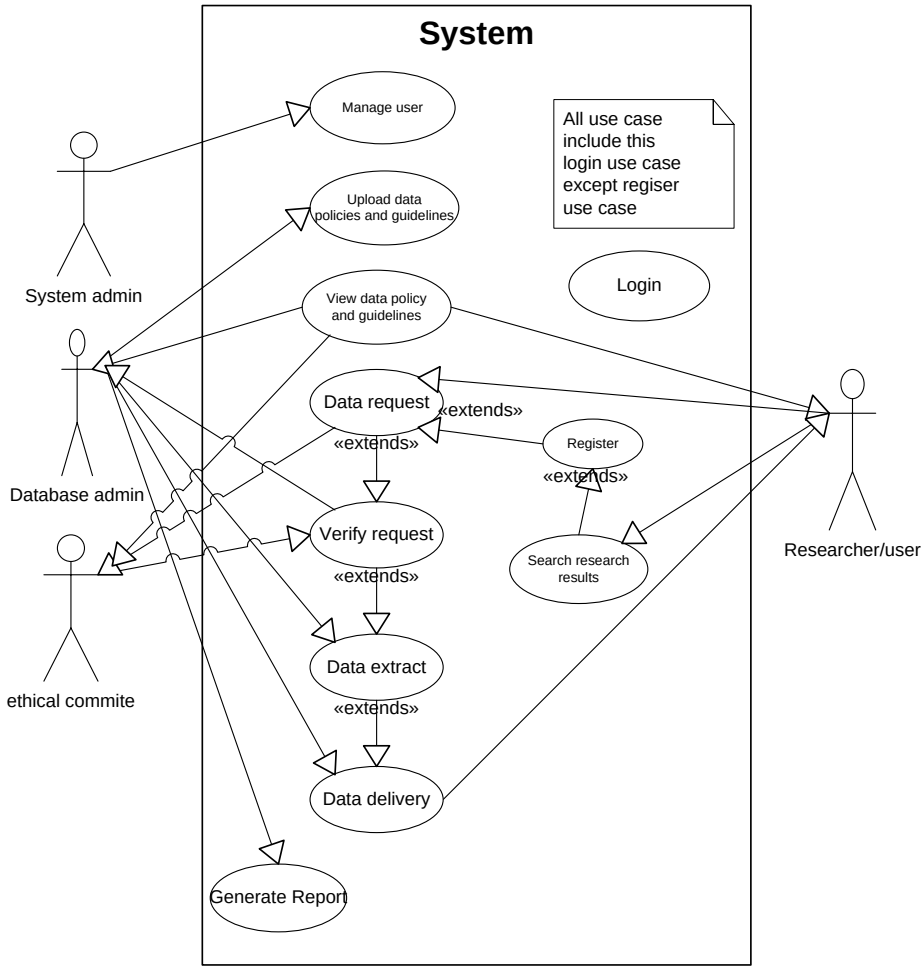


Figure 4: Use case Diagram of the proposed system

Use cases of the system

- Upload
- View
- Register
- Searching
- Login
- Request data
- Verify request
- Import/download data
- Deliver data
- Generate Report

Table 5: Upload /update data

Id	EPHI-001
Name	Upload /update document or data
Description	Administrator enable to upload data policies and guidelines to the system
Primary Actor (s)	Administrators.
Pre condition	None
Post condition	EPHI-006
Basic Scenario	1. The administrator wants to upload data 2. The administrator login to the system 3. The system displays the administrator page 4. The administrator selects the upload option. 5. The system display upload dialog box. 6. The administrator select the document and click upload 7. The system processed uploading and the document loaded and display in the menu. 8. The use case ends
Alternate scenario	7A.1 if the system processed uploading and display a message "failed, It is big size 7A.2 if the processed upload and display a message " failed, try again"

Table 6: Search research results system use case

Id	EPHI-002
Name	Search
Description	User enables to search research results, using title, year and study type as key wards for searching.
Primary Actor (s)	All users.
Pre condition	None
Post condition	None
Basic Scenario	1. The user wants to search research info 2. The user browse the system home page 3. The system displays the home page 4. The user uses the 'Search' option and writes the key words on the search box and press search key. 5. The system searches all the items matching with the given keywords and displays a list of the matching items. 6. The user selects the one he/she want review 7. The use case ends
Alternative Scenario	5. A. If The search not found matching items and the system display a message "doesn't exist"

Table 7: View guidelines system use case

Id	EPHI-003
Name	View data
Description	Users enable to access available research results or documents
Primary Actor(s)	All users
Pre condition	None
Post condition	None
Basic Scenario	1. The user wants to read document 2. The user browse the home page 3. The system display the home page 4. The user select one of the option from the menu 5. The system display the page 6. The user click on the document 7. The system display the required document 8. The use case ends
Alternative Scenario	7A1. if the browser unable continue process to open the document and the system can display error message like, failed the process , update your browser 7A2. if the required document unable to open compatibility problem or the source document not available can display the system.

Table 8: User registration system use case

ID	EPHI-004
Name	Registration
Description	User's information capture in the system
Primary Actor(s)	All actors except the anonyms user
Pre condition	None
Post condition	User registered
Basic Scenario	1. A person wants to register in this system 2. A person browse the home page 3. The system displays the home page 4. A person select register option 5. The system display registration form 6. A person fills the registration form 7. The system checks a person inserts the necessary information, or whether a Person registered before or a person uses right data format. 8. A person clicks on the submit button 9. The system save and store the person information into the database and display confirmation 10. The use case ends
Alternative Scenario	8. A1. If wrong data format is inserted the system will notify to correct the information 8. A2. If skip the required filed the system display a message to filled the field 8. A3 if a person's information found in the database, the system display a message "you are registered" and reject

Table 9: user login system page

ID	EPHI-005
Name	Login
Description	Users enable login into the system
Primary Actor(s)	Users who are registered in the system
Pre condition	Registered
Post condition	User logged to the system
Basic Scenario	1. The user wants login into the system 2. The user browse of the system 3. The system display the home page 4. The user selects the login options 5. The system display the login page 6. The user inserts username and password. 7. The system authenticates the username and password of the user. 8. The system brings the required interface that is allowed for that user 9. The use case ends
Alternative Scenario	5. A1. if the username or password is incorrect the system displays a message "invalid username or invalid password or invalid username and password " 5. A2. if the user does not use the correct password the system will lock sown after five trial

Table 10: Data request system use case

ID	EPHI-006
Name	Data request
Description	Users form different health sectors to have research data access
Primary actor(s)	Users
Pre condition	EPHI-006
Post condition	
Basic Scenario	1. A user wants to request data from EPHI 2. The user browse the system home page 3. The system displays the home page 4. A user login into the system "EPHI-006" 5. The system display user page 6. A user select data request form 7. The system display data request interface 8. A user fill the information and submit 9. The system checks a person inserts the necessary information or uses the right data format. 10.The system saves and stores the person information into the database and display a message accept request 11.The use case ends
Alternative Scenario	10. A1. if wrong data format is inserted the system will notify to correct the information 10. A2. If the person skip the required filed the system display a message to be filled the field

Table 11: Verify request system use case

Id	EPHI-007
Name	Verify request
Description	The research and ethics committee verify the requested data access before allowing user to grant access to research data
Primary Actor(s)	Research and ethics committee
Pre condition	EPHI-006
Post condition	
Basic Scenario	1.The research and ethics committee wants to verify request 2. The research and ethics committee login into the system using " EPHI-006" 3. The system display research and ethics committee page 4. The research and ethics committee view the request and review 5. The research and ethics committee approved the request(the system mark (flag) the request as approved) 6. The use case ends
Alternative Scenario	5.A. if there is error message from the system , user try again

Table 12: Data extract system use case

Id	EPHI-008
Name	Data extract
Description	The administrator enables to extract data difference sources
Primary actor (s)	Database administrator
Pre condition	EPHI-006
Post condition	
basic Scenario	1.The administrator wants to data extract 2.The administrator login "EPHI-006 3.The system display the data administrator page 4.The administrator view the request 5. The system display the required data information 6. The administrator import the required data 7. The system processed and get the result 8. The administrator select request query 9. The system display the result 10. The data administrator checks the data is required and save a temporary media 11. The use case ends
Alternate Scenario	7 A1. if the system unable continue processed the function and display a message, failed the process, there is an error 10A1. If the data is not hold the required information. the data administrator try again the function

Table 13: Generate report system use case

ID	EPHI-9
Name	Generate Report
Description	Database administrator enables to generation report
Primary Actor(s)	Database administrator
Pre condition	EPHI-006
Post condition	
Basic Scenario	1.The data admin wants to generate report 2.The data administrator login "EPHI-006 3.The system displays data administrator page 4. The administrator select report generates option 5. The system display the result 6. The administrator save the result 7. The use case ends
Alternative Scenario	5. A1. if there is no data match request report 5. A2 The system prompts no matching data found 5. B1 if the administrator wants to have the report in different format 5. B2. The administrator click on the export report button and Select the file type of the report 5. B3 the report viewer can save the report on his/her pc

4.6.4.3 Sub system description

To be understandable the whole system can be broken down into the subsystems. These subsystems group those functionalities which have the same purpose into one group.

Interface sub-system

The designed health research data repository information system has a total of ten basic subsystems interfaces under the overall system functionality.

Manage user account: enables the system admin to create user account and set privileges for different system users; - moreover the system admin can do enable or disable users in the system.

Upload: is used to enables the database admin to provide the data policies, guidelines, and other research resources in the system

Searching: is used to enable system user to get research resources from the system

View: this interface is used to enables users to access the research resources

Register: This user interface enables capturing user information on the system and gets confirmation, capturing information about research result, source data, research document and statistical software

Login: this user interface is used to validate and authorize users to access the user page of the system.

Request data: is used to enables users to fill request data form and submit to the system.

Verify request: is used to enables the ethical and research committee to open and approve users Data request by viewing the data request form in the system.

Extract data: is used to enables the database admin to import and extract research data through the system.

Deliver data: is used to enable the database admin to fill data deliver form in the data sharing process conformity.

4.6.4.4. Class modeling.

Class modeling is an essential part of UML and shows the object classes in the system with their association between them. Each class may have to have unique name, some knowledge of its associated class (attribute) and the actions that class performs (operations) are found together in a single block. These classes are capture information about public health research in developing an object-oriented system modeling that can be used as to provide a general overview of objects and their interactions in the designed repository system as show in the figure 5 below.

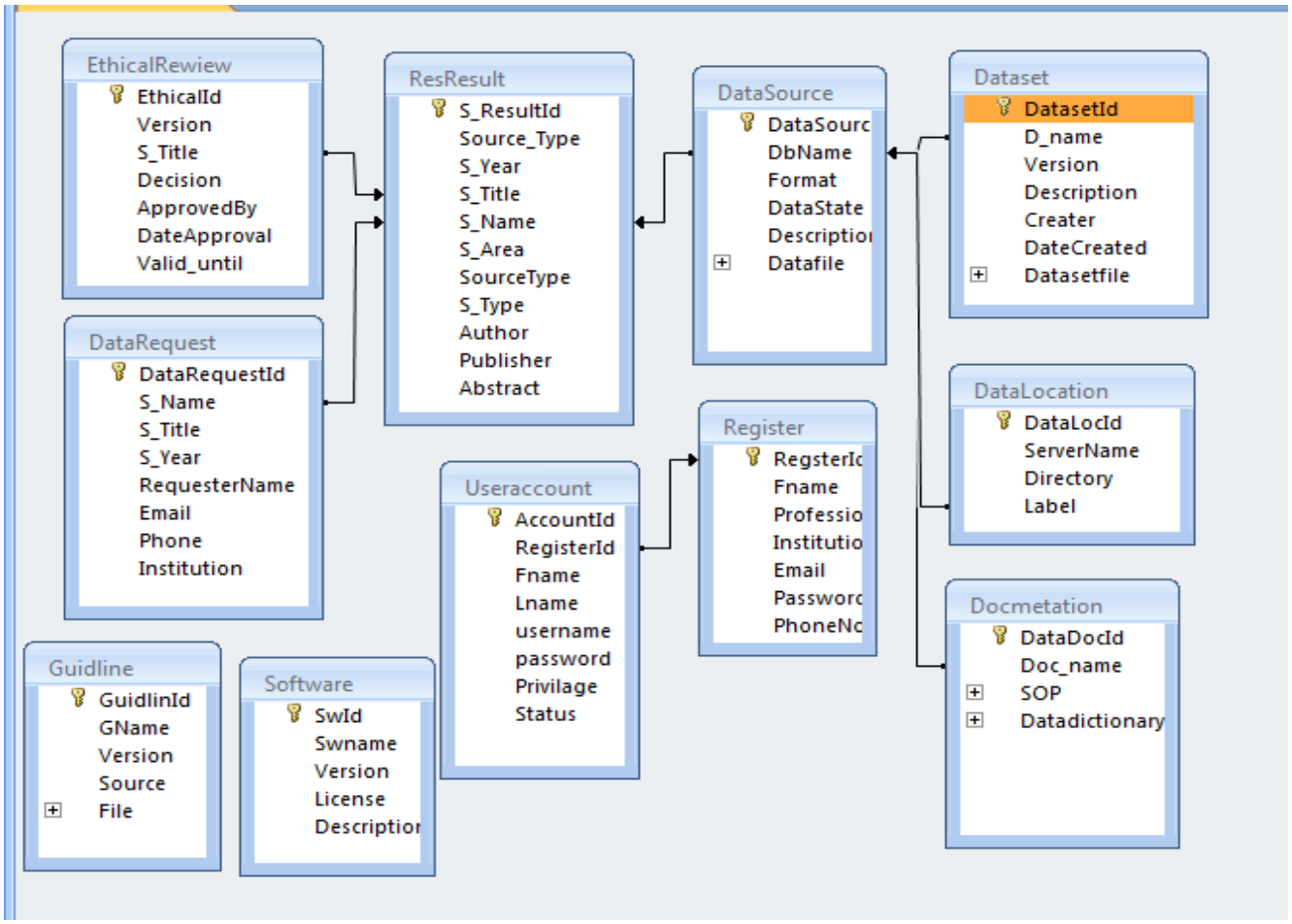


Figure 5: Class diagram of the system

4.6.4.5 Sequence modeling

One of UML modeling technique is the sequence dynamically model that supports view of developing systems and shows the explicit sequence of messages are passed between objects in a defined interaction. Objects are indicated by annotated arrows and the rectangle on the dotted lines indicates the lifeline of the object concerned (Dennis, 2012). Sequence diagram employs representation for the existing objects and sequence interaction a particular use case that allows different kinds of interaction to be form in the repository system. Figure 6-11 present the sequence diagram of the proposed system.

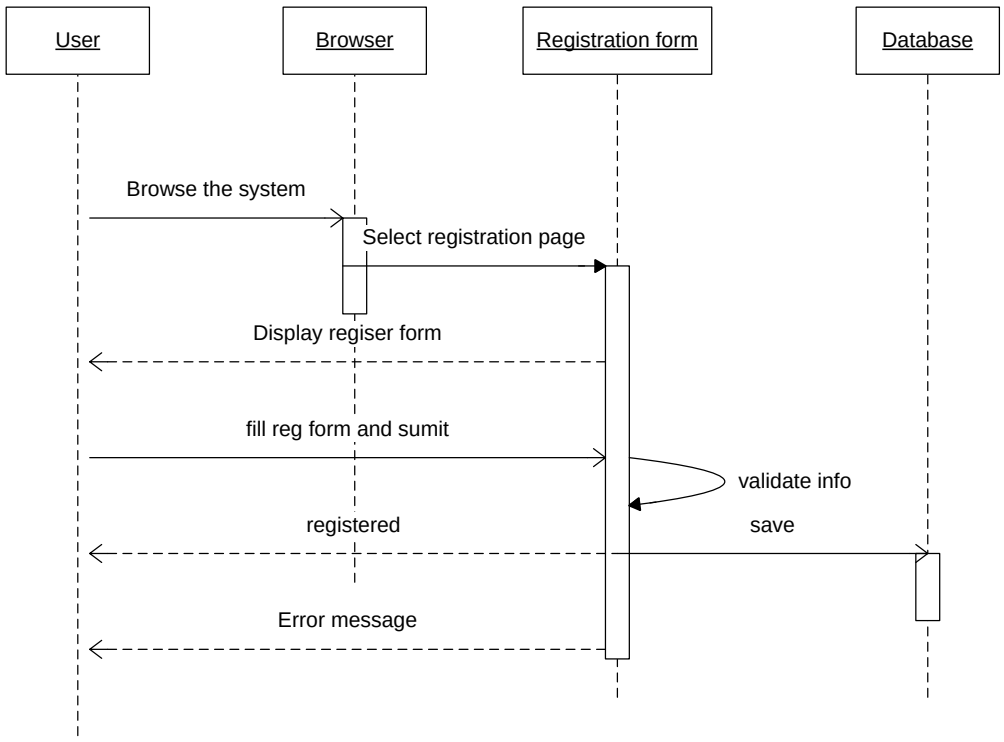


Figure 6: Sequence diagram user registration to the system

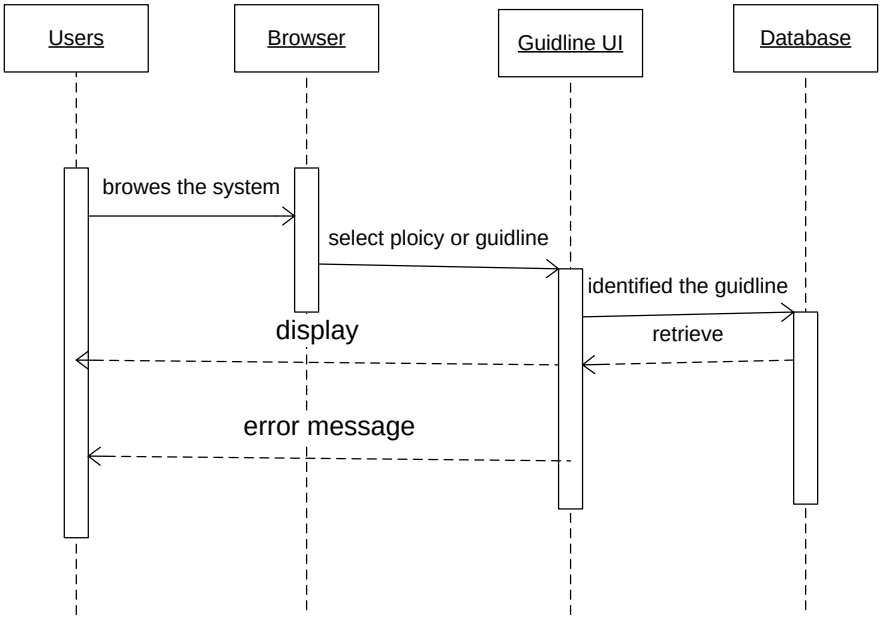


Figure 7: Sequence diagram for view guidelines system

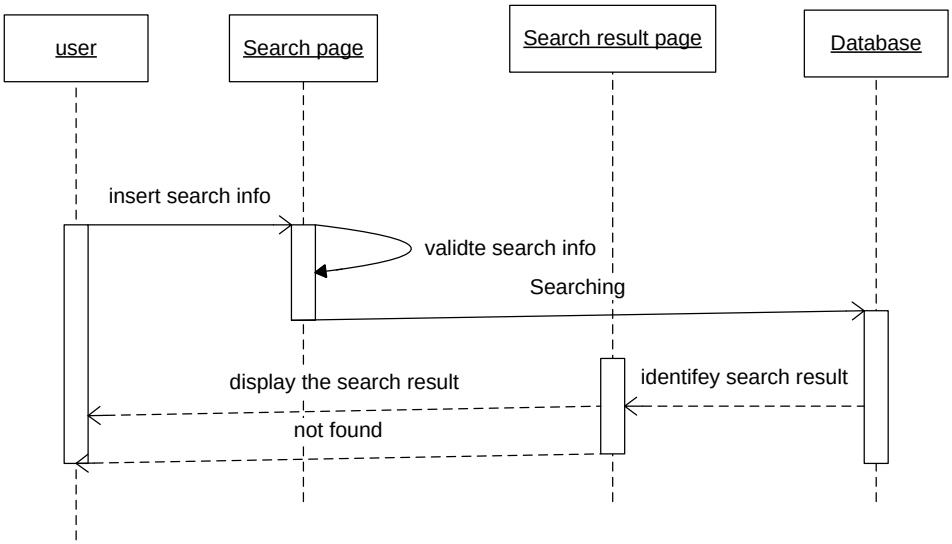


Figure 8: Sequence diagram search for research resources system

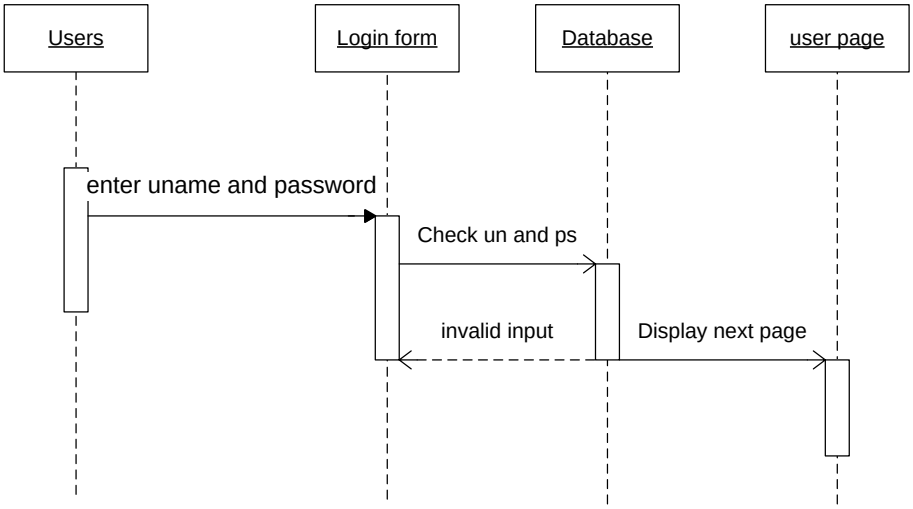


Figure 9: Sequence diagram for user login to the system

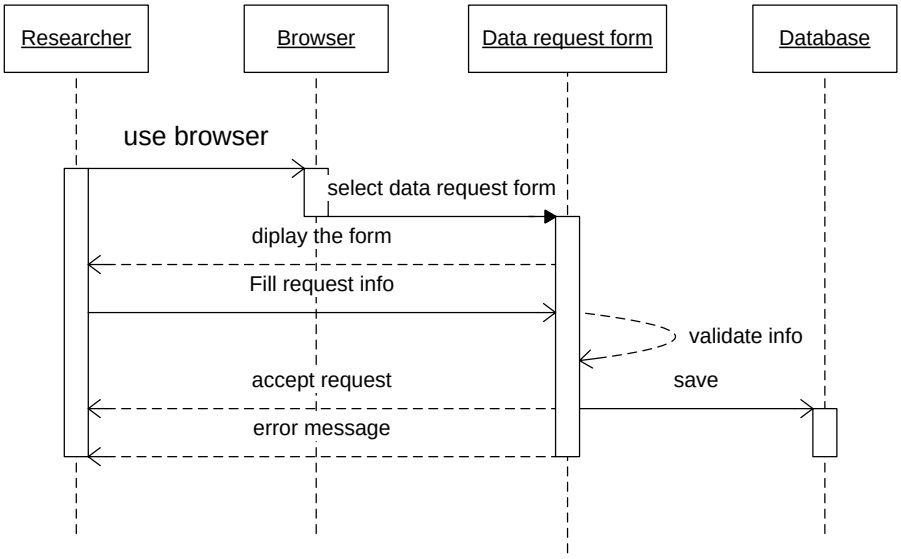


Figure 10: Sequence diagram for user data request system

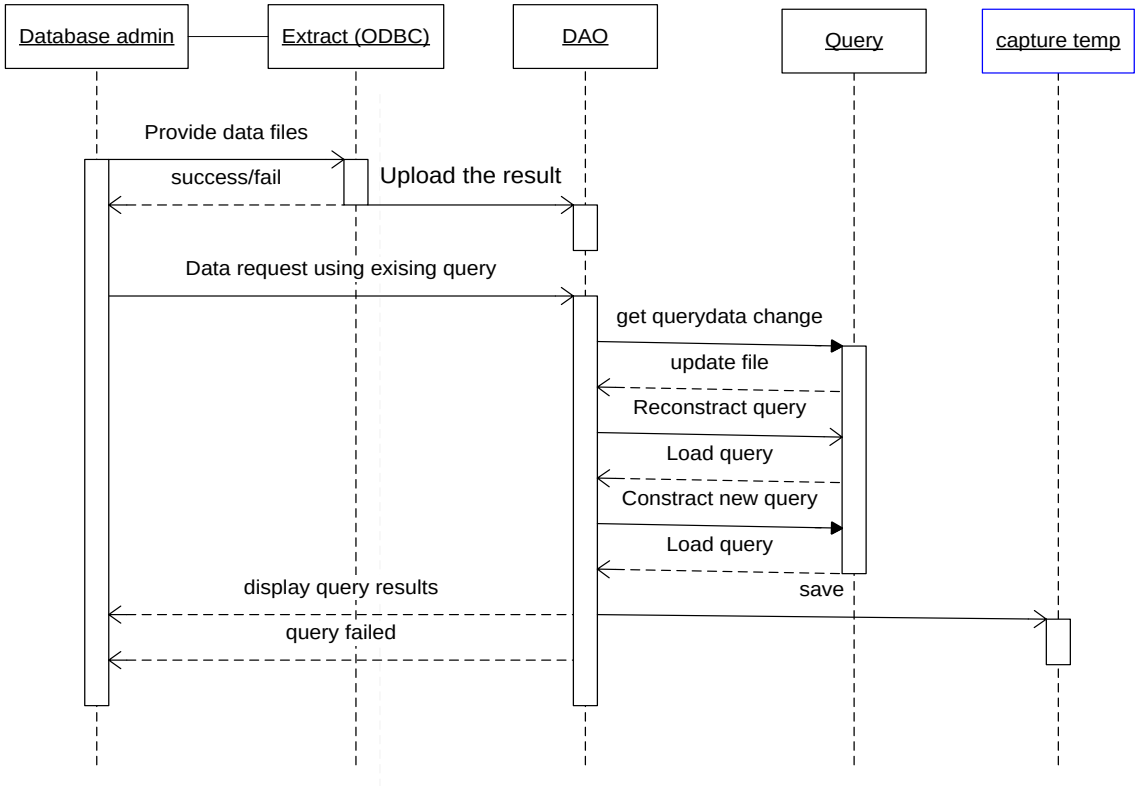


Figure 11: Sequence diagram data extraction system

4.6.4.6 State diagram

State diagram describes the behavior of a single object in response to a series of events in a system. This UML diagram models the dynamic flow of control from state to state of a particular object within a system (Jacobson, 2006). For instance, to get research data from the EPHI, users have to pass process starting from data request to delivery different state in this repository system. This process passes series events (data request, verify request, searching data sources, data extract and delivery) to accomplished as shown in the figure 12 below.

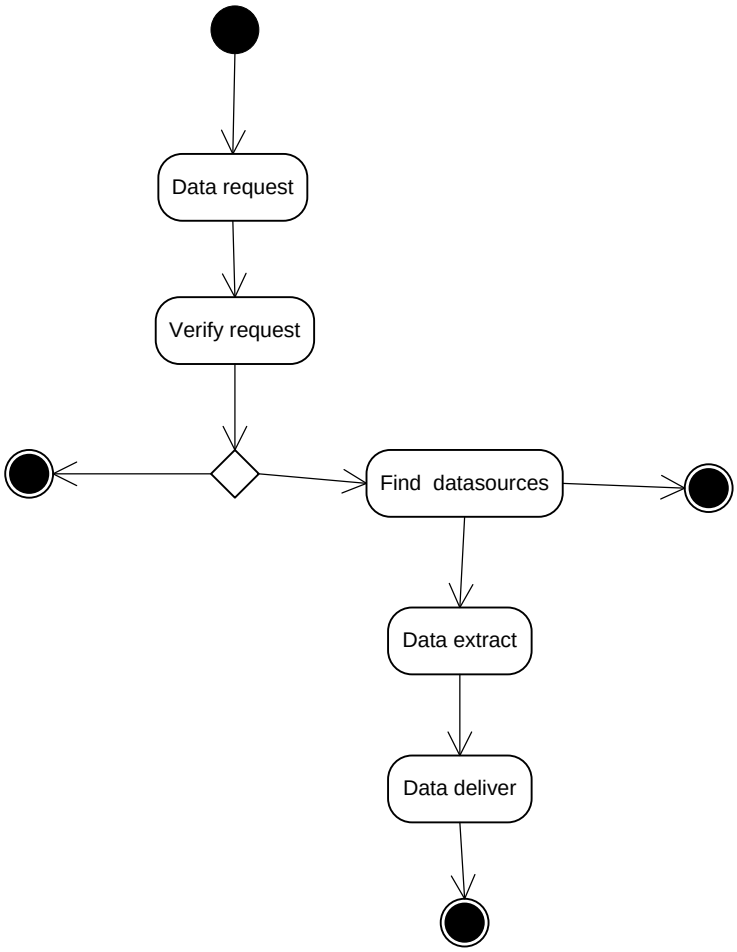


Figure 12: State diagram for data request processes

4.4.5. System deployment Architecture

Deployment diagrams show how the system software is deployed on a particular hardware configuration. The implemented system has to run on hardware, meet its nonfunctional requirements such as performance and scalability. Deployment architecture shows the structure of the runtime system: which component implementations run on which processors and how the hardware is configured to provide necessary resource (Jackson, 1999).

The public health research processes a complex heath data and involves various uses from different place of health sector. To addressing those geographically differing and information needs, the proposed system is built in 3-tier application architecture. This layering is logical meaning that the layers could be deployed within a single or two machines. However, in the case of large scale applications it also becomes physical separation and each logical layer is deployed separately in order to increase performance (Grozev, 2013).

The first tier (Front end) of the system

The presentation layer implements the user interface which the end user interacts with. It can be either a separate application or web interface accessed through a web browser. Either way it is executed on the end user's machine and thus we do not take it into account in the servers' performance. User registration, upload ,view, login , searching are common interface of the ystem.

The middle tier of the system

The middle or business layer implements the essential application logic and rules in which the system has to perform. This includes all application specific business functionalities. Traditionally, it is executed in one or several application servers which communicate with the presentation and the data. This layer acts as a bridge between the user and the database. Various queries that are found in this proposed system are found on this layer.

The third tier (back end) of the system

This is data layer that facilitates access to the persistent storage. It is independent from the other two layers, as the persistent data often survive the applications that use it and it can even serve more than

one application. The data layer is executed in one or several database servers and stores and provides the different data that are required for the full functionalities of the system. In this layer the different data that are captured via user registration, research result data, data source and other research resources data are stored and used for retrieval. System architecture shows in Figure 13 below.

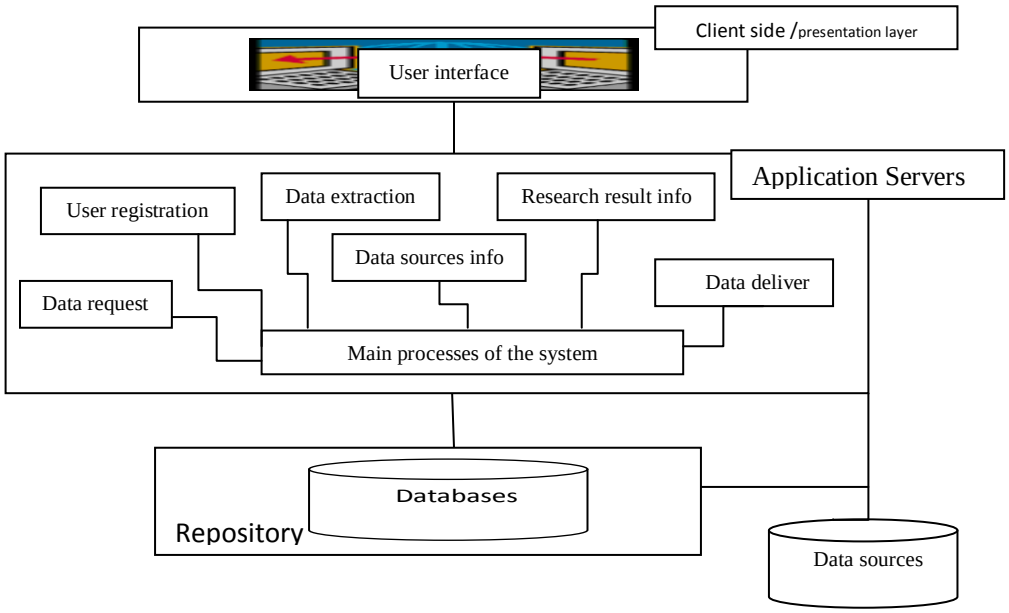


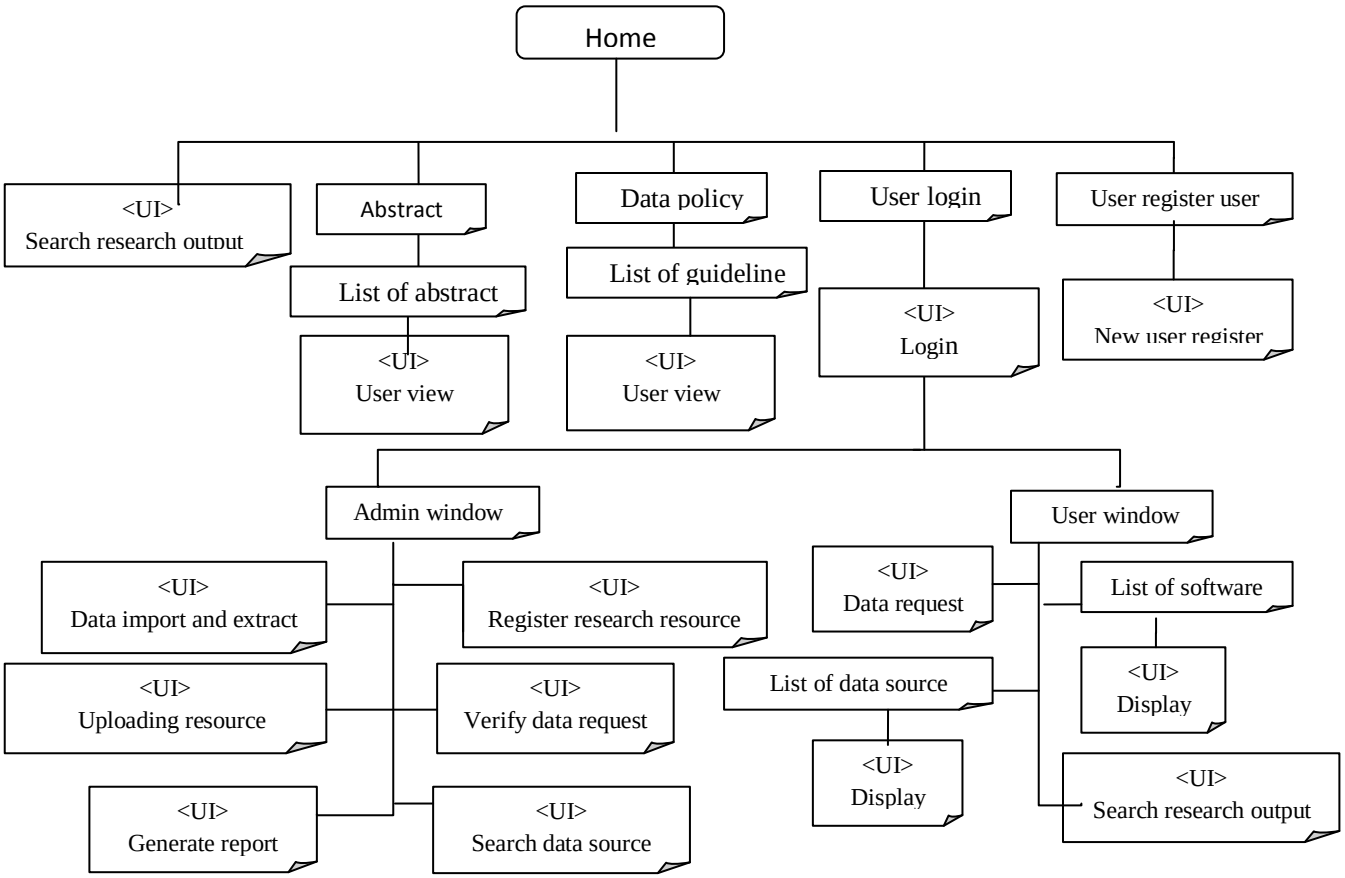
Figure 13: Deployment diagrams for the proposed system

4.4.6 User interface prototype

User interface is the part of the system with which the users interact. It includes the screen displays that provide navigation through the system, the screens and forms that capture data, and the reports that the system produces (whether on paper, on the Web, or via some other media) (Dennis, 2012).

User interface prototyping is a testing and evaluation approach, which is a crucial component of User-Centered Design (UCD) methodology. It is considered as an excellent approach for facilitating communication between the designers and users. Prototyping is not only helps to visualize design concepts in an interactive way, but also supports expressing new requirements and expectations towards a prospective system. Usability testing of user interface prototypes validates user requirements and functionality of the prototype under evaluation (Weichbroth, 2015). User interface flow diagram of the proposed system is presented in Figure 14 below.

Figure 14: User interface flow diagram for the proposed system



User Interfaces of the integrated health research data repository web based system

Figure 15: Home page of the system

This home page of the system provides list of research abstract, data polices and guidelines with various social media links easy accessible for users. This page enables user to search research result and redirect to login/register page that user to register and have got login access privileged further functionality of the system.

Figure 16: List of research abstracts interface

This interface uses to display list of public health research abstracts.

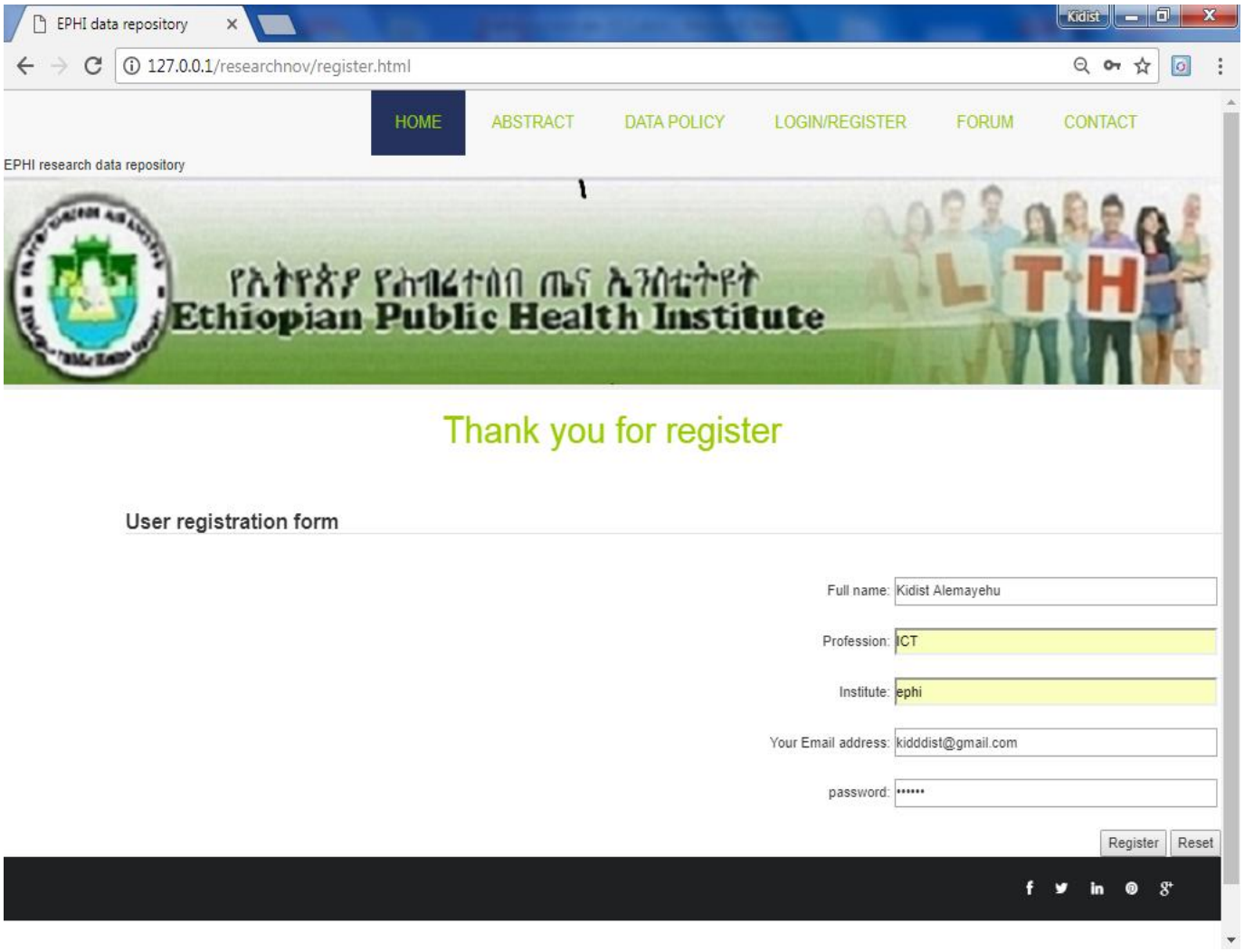


Figure 17: New user registration form

If users wants member of the system or need more information about research data source, by register user get user account to have access.

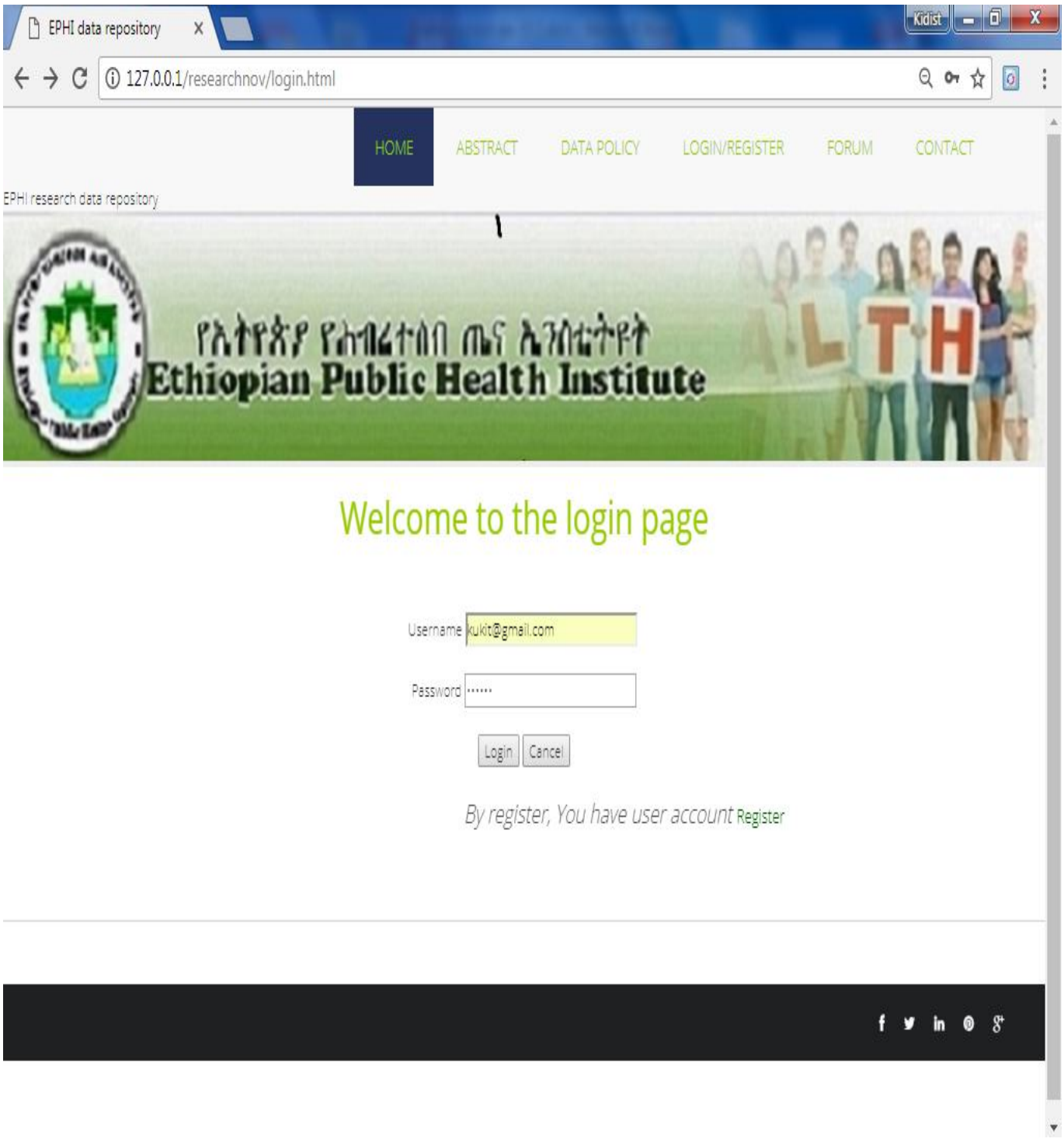


Figure 18: Login interface

This user interface is a gate way to access user page of the system. Each user feeds his/her right user name and password to the given login form, can logs into the system.

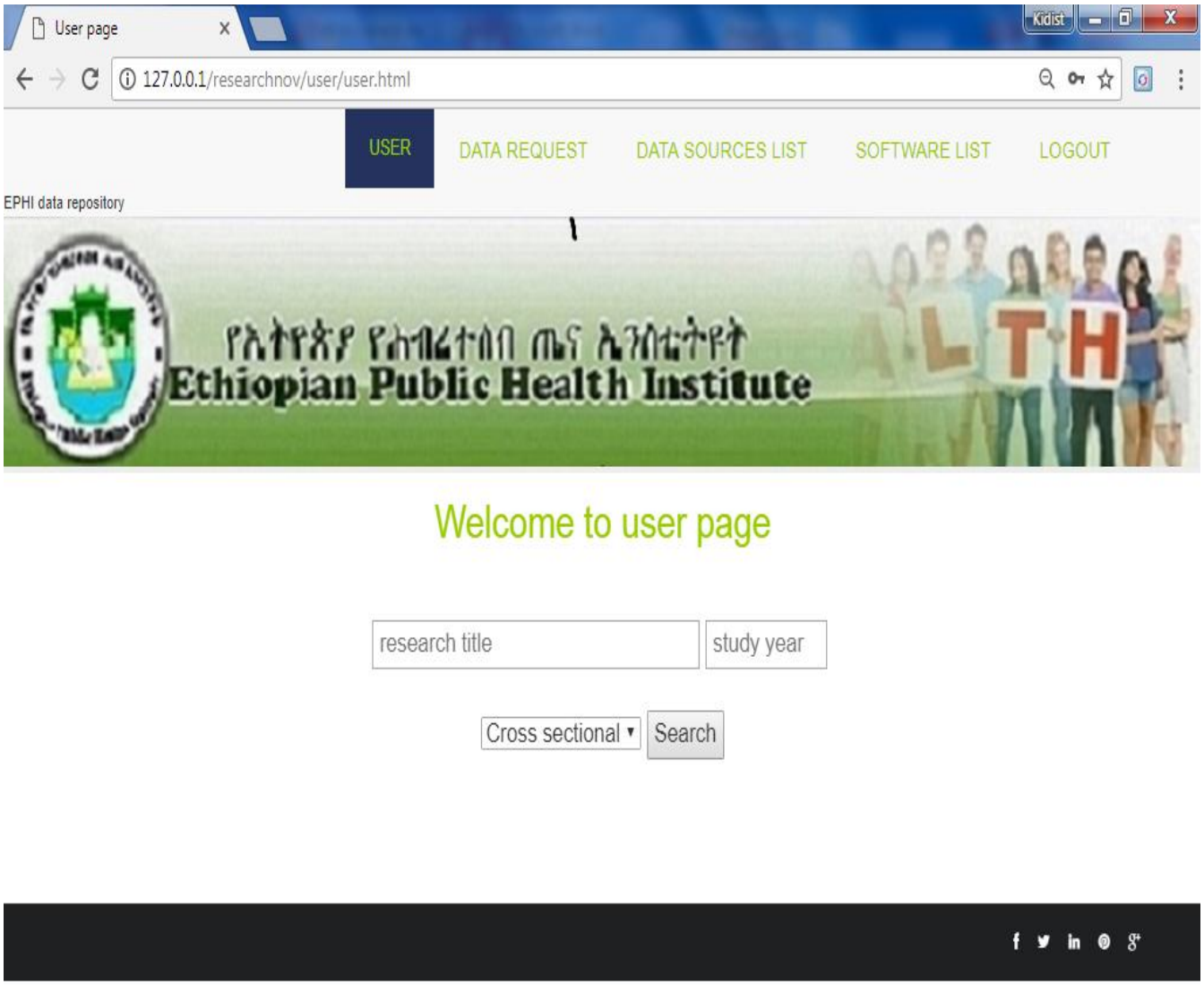


Figure 19: User interface page

This user interface provide additional information for registered user which is not find in the home page, data request form, display list of available research data sources and display list of statistical soft-wares, are helps to user informed about the available research resources.

Figure 20: List of data source user interface

This interface display research data sources with their short descriptions which is name of the data base, the data state (raw/aggregate) of stored and the application of the database.

Figure 21: List of software user interface

This interface use to display list of statistical software that are commonly used for research purpose.

Figure 22: Data request user interface

When users want to get research data from the institute, fill and send this form to ethical committee to approve their request.

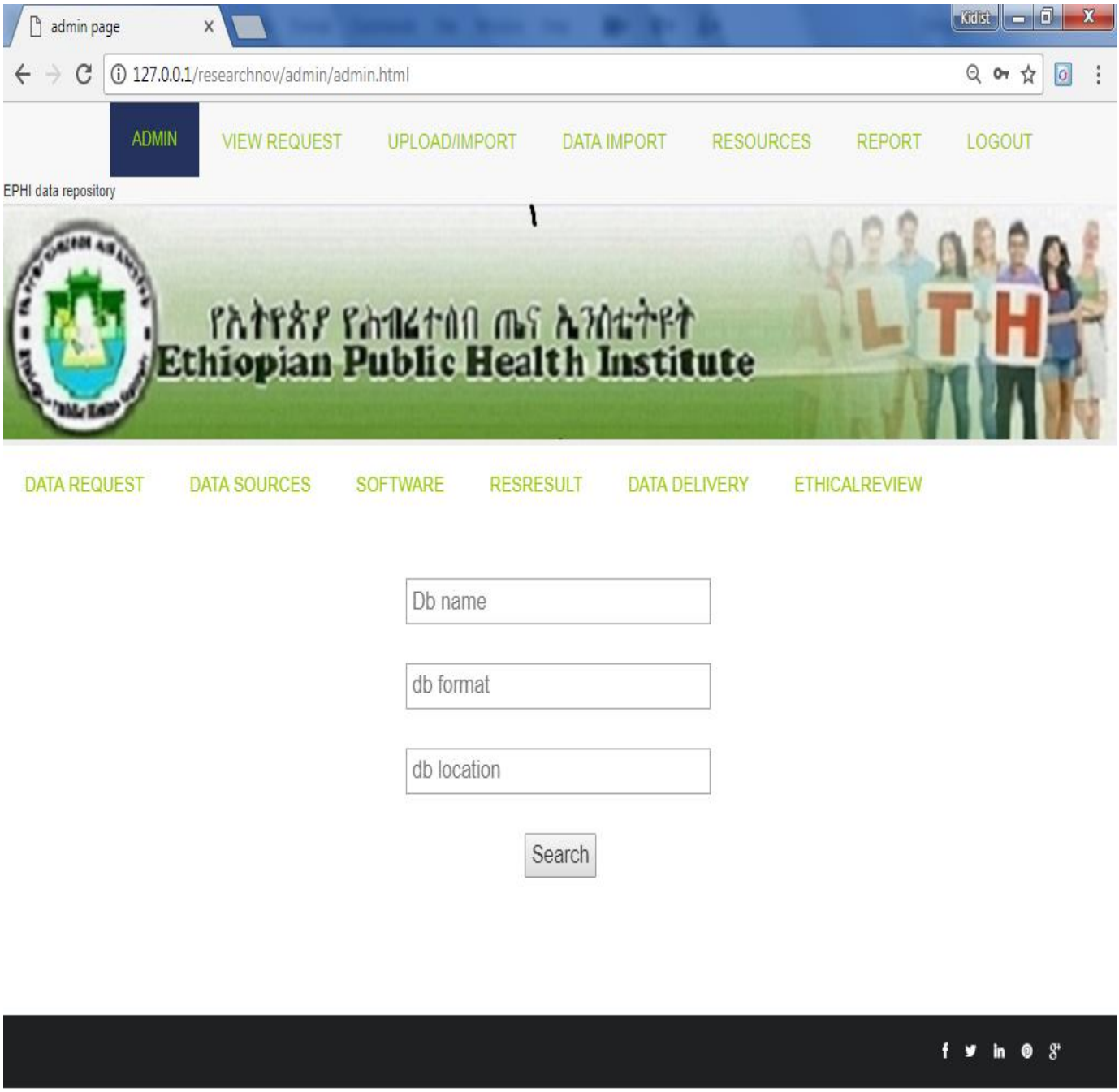


Figure 23: Administrative user interface

This page holds different part of the system which managed by data administrator, like uploading research resources, import data source for data sharing purpose and capturing research resources information.

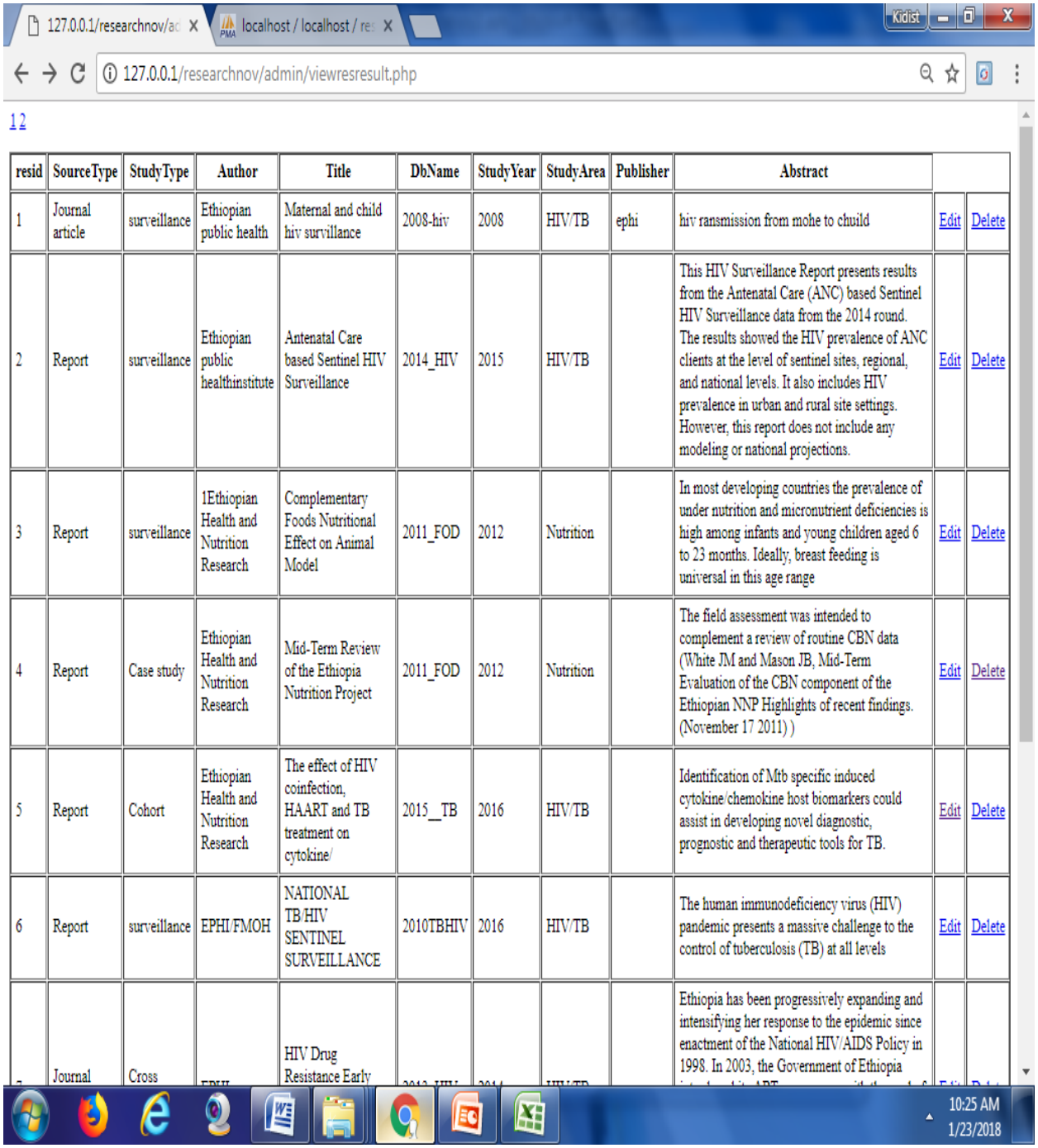


Figure 24: View list of research result user interface

This page use to view the recorded research result available in system.

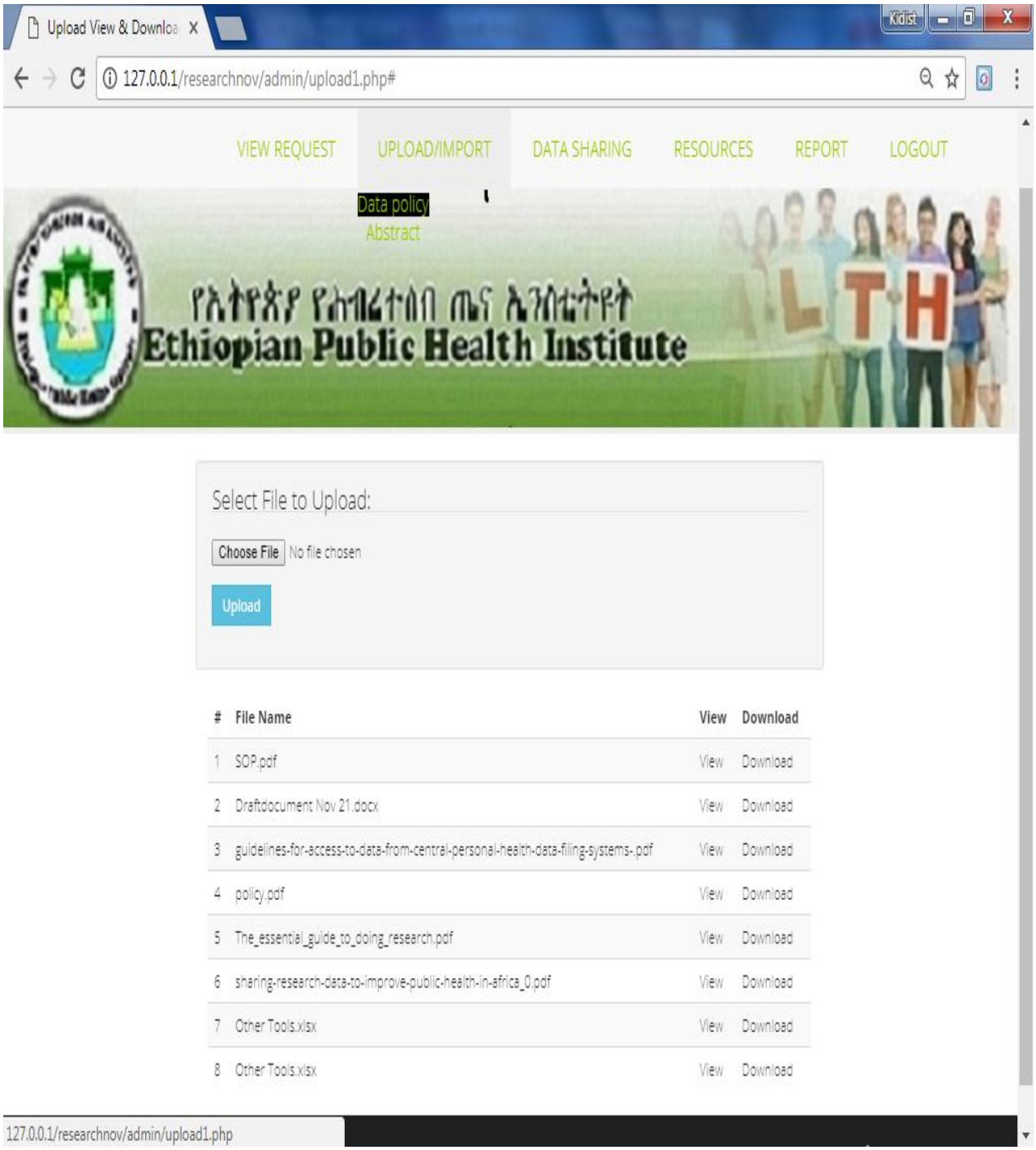


Figure 25: Upload/download user interface

This interface use to upload and download guidelines in the system by data administrator

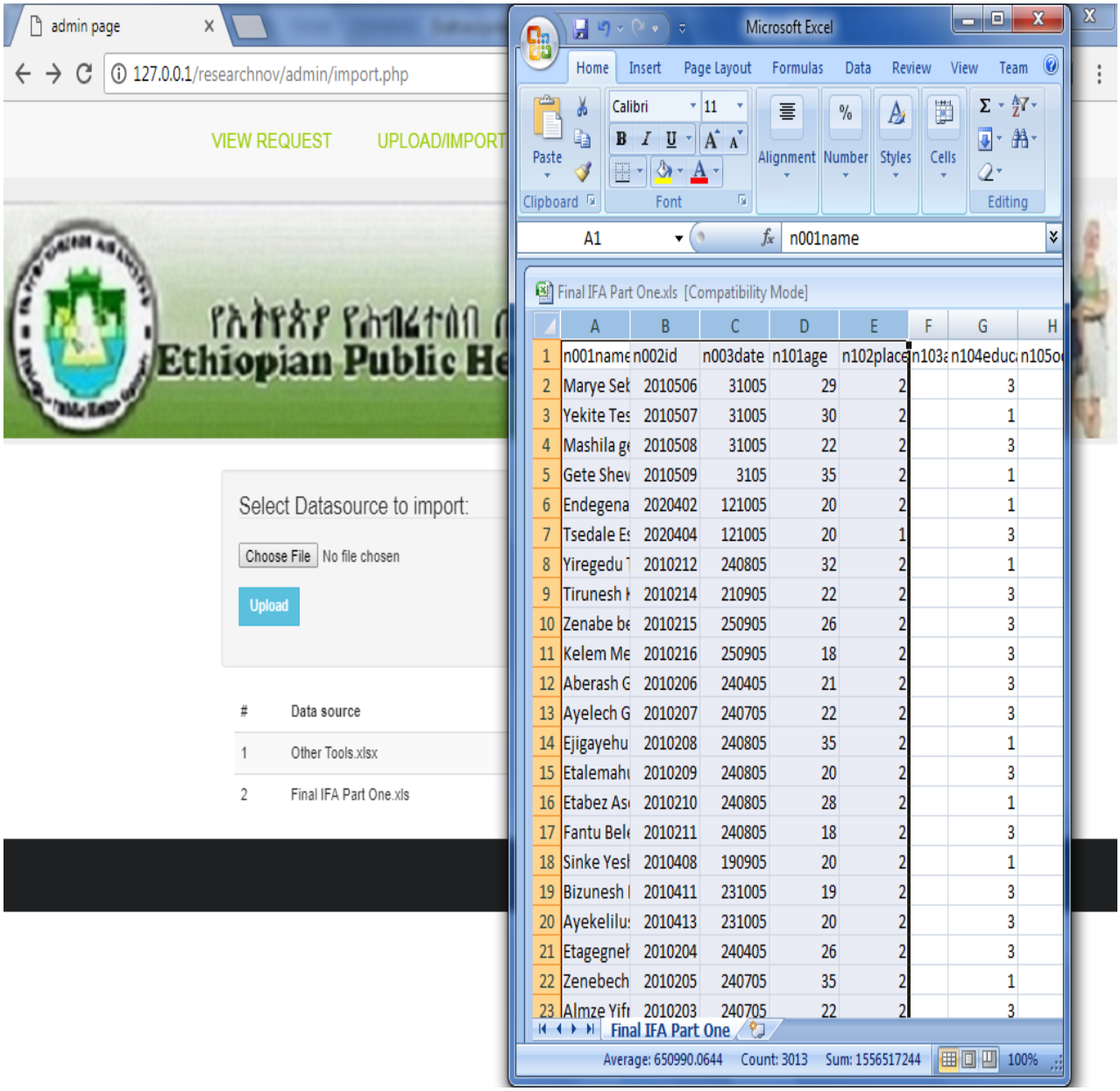


Figure 26: Upload/import data source user interface

This interface uses the purpose of data sharing that import the requested data source and select the variables.

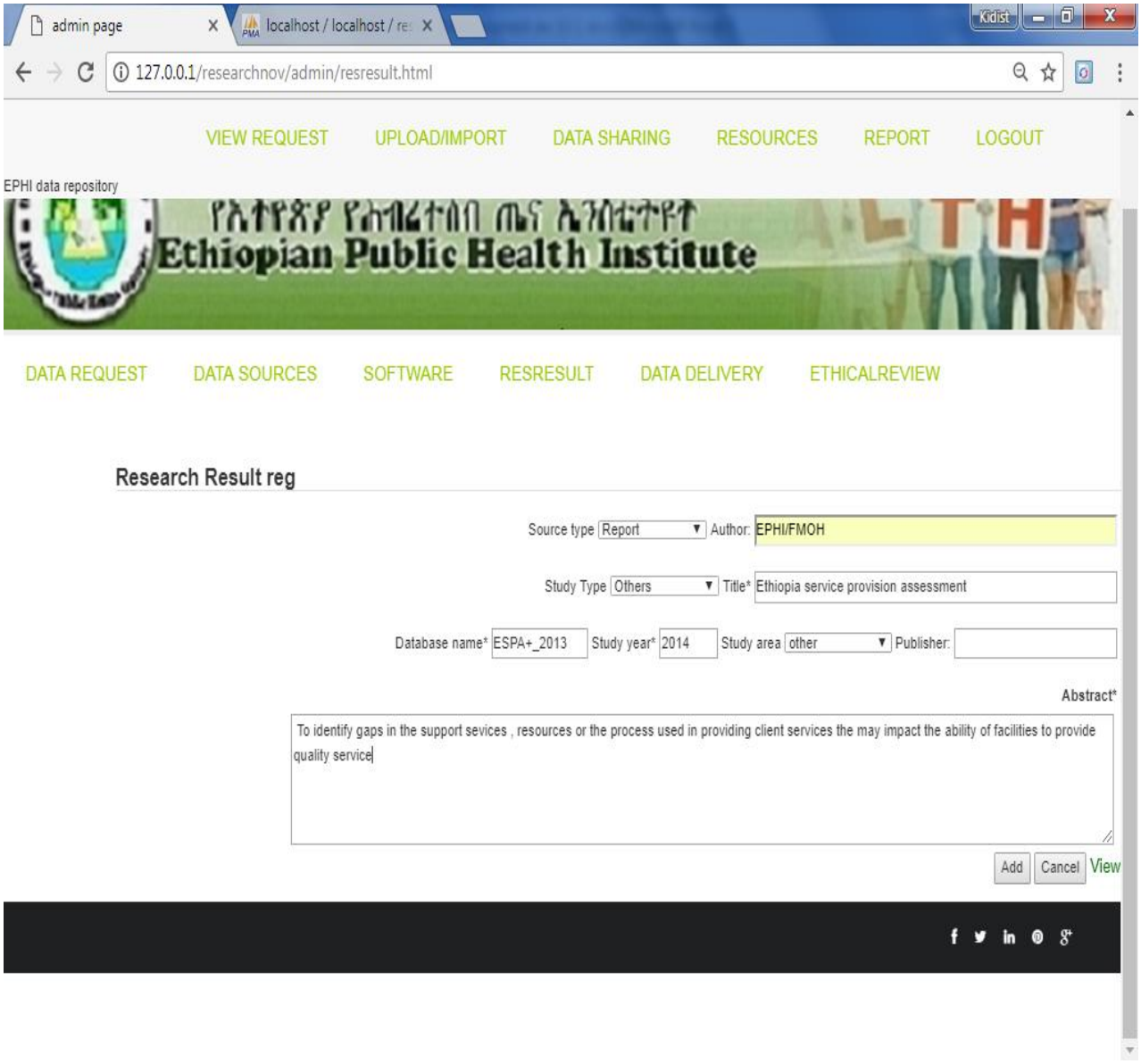


Figure 27: Research result registration form user interface

This user interface is used to capture the basic and necessary data about research results.

4.5. Usability test

Usability is the degree of user reaction towards a computer system to achieve specified goals efficiently while promoting feelings of satisfaction in a given context of use with the usability principle. The reaction of the users to an interface is one of the key things which are studied in human computer interactions (Oluwatosin, 2016).

The interface evaluation consists of methodologies for measuring the usability aspects of a system's user interface and identifying specific problems to improve the interface design. Interface design is subjective; there are no formulas that assurance a great user interface (Hartson HR, 2003). During the interface evaluation of system users should be representative of the typical user of that specific application and the number should be of optimal size for the anticipated evaluation results. The tasks should represent a whole user task and will allow assessment of the interface for consistency between user's conceptual models of the tasks. The usability tests should be performed on real products or working prototypes (ISO 9241-11, 2011).

Heuristic inspection method was applied in this repository system. Inspection the designing system of prototype interface which asks usability practitioners and other stakeholders to evaluate the different part of a system based on a set of design principles in easy and cost effective methods. The evaluation of the interface is one of the key aspects of successful software design and development. The evaluation of the interface web based health research data repository system used questionnaires to evaluate the interface for the prototype developed. Questionnaire method that has been adopted is used, for it is cost effective and using Likert scale evaluation methods (Kotari, 2004).

Table 14: User interface evolution result

No	Questioners	Strongly disagree	Disagree	Neutral	Agree	Strongly agree
1.	It was easily manageable options of the system			2	4	4
2.	It was clear to access the different options				2	8
3.	The system provided adequate functions			1	4	5.
4.	It was clear on using the different functions of the system				1	9
5.	The system flexible in allowing the user to choose options				2	8
6.	The system handle errors in user actions		5			5
7.	Overall, the interface was user friendly and attractive					10
	Average result	0	7.1%	4.29	18.57%	70%

The result of the user interface evaluation has been shown in above table. The values of the questionnaire responses are taken proportion in the five categories. 88.57 % of the respondents agreed with the system prototype have user acceptable interface design.

CHAPTER FIVE

Conclusion and Recommendation

5.1. Conclusion

EPHI is a partner institutional organization under the FMoH mandated to conduct public health research in the country. As research institute, data collection is done in different time for various public health researches and the data are stored. Hence from time to time increasing amount data and becoming challenge to handle in the existing system. Accordingly the respondents of the study, the existing system problems regarding in their research process and related with data storage utility, lack of information about available data in the storage, lack of disintegration research resources, due to disorganized data sources difficult of traceability, lack of systematic data sharing and accessibility that to promote secondary use. This research project can contribute to fill the gap and improve the institute research process performance and storage utilization.

Based on the requirement gathering done on the basic information system components; the need for information capturing and uploading research resources (research results, data sources, research guidelines and polices, research software and other research information) and extracting data for further research and there is informal data sharing were identified. According to the response by the respondents, there is no way of optimal storing to integrate the research finding with its research data source that helps to easily accessing and treacing the available research resource for potential users. Besides these, useful documentations (guidelines, data polices, data sharing agreement and SOP) are not regularly practical for every research which conducted in the institute. This research project by assessing the problems, from the group opinions of the researchers and using policy guide lines developed the necessary requirements of the system to be designed.

After having the essential requirements, the design and analysis for the system was done. This step involved the stakeholders of the system. The functional requirements were gathered and developed to address the major business requirements of the system that the stakeholders need from the system. These functional requirements included the user registration, verify request, reporting, data requesting, ethically clearance, uploading and extracting. Beside the functional requirements, the non-functional requirements which are required to keep the system running like security, scalability, interoperability and performance requirements were also identified

The design of the system was accomplished based on the requirements analyses. The design of the system applied objects oriented analyses and design methodology. In order to transcribe the analysis into a design, UML was used. Among the different types of the UML modeling tools (use case, sequence, class diagram and state diagram) were used in order to implement three tired system architecture. These can enable a secure and networked web based system.

Through all the inputs from the requirement analysis and design, a prototype to the system was designed to show the working model of the system. The user interface prototype consisted of the display, different functionalities and navigational need of the users. The satisfaction of the users on the developed prototype can be achieved by evaluating the prototype. Finally the web based integrated health research data repository system addresses one of the core business processes of research information handling needs of the EPHI and improves the overall research process involved in the institute public health research service.

5.2 Recommendation

In this study, an attempt to health research data repository system and develop a system that will benefit public health researchers involved in the process. Accordingly the finding in the study the following recommendations were drawn and these recommendations should be taken into consideration by EPHI, and the other stakeholders which are directly and indirectly attached to the system.

.

Ethiopian Public Health Institute

- This system can be the baseline of the establishment for national public health research data repository system that can be depository of every public health data in the country.
- Work on the implementation of the system and provide the necessary system support.
- Encourage users who inquire research data access to use the system for their request which is increasing usability.

Researcher / research group (target user)

- Understand the possible impact in the research environment if their research paper browsed by others easily in usability of the system which is useful to fill research gaps and practices the group is using to get to work faster.

.

EPHI's data administrators

- As data administrator should working on the administrative tools provided by the system that strengthen technically supportive the research process and protect from viewable to anyone unauthorized by accident and without explicit permission. In this idea each researcher would agree and have trust on the system

Researchers/students

- To continue the project and work on the development and implementation of the system and with incorporate the remaining part of the system.

Reference

Ambler, S. W. (2005). A Manager's Introduction to The Rational Unified Process (RUP).

Andreescu, A. I. (2012). Perspectives on the Role of Business Rules in Database Design. Database Systems.

Athina, L. (2010). Web-Base Application in Healthcare and biomedicine. London.

Austin, C. (2015). Research Data Repositories: Review of Current Features, Gap Analysis, and Recommendations for Minimum Requirements.

Bernstein', P. A. (1994). An Overview of Repository Technology. Proceedings of the 20th VLDB Conference, Santiago Chile.

Bihammar, P. (2007). Melbourne Health Realizes Research Efficiency with Information Integration.

Booch, G. (1999). Unified Modeling Language User Guide. Addison Wesley.

Brody, T. (2007, July/August). Key Principles in Assessing Repository Models. 2007 Issue of D-Lib Magazine

Brown BB, F. L. (2016). 6 Steps for Implementing Successful Performance Improvement Initiatives in Healthcare.

Bugeia, P. (2013). 4A Data Management Acquiring, Acting-on, Archiving & Advertising research data at the University of Western Sydney.

Castro, D. (2009). The Role of Information Technology in Medical Research. The Information Technology and Innovation Foundation .

Çeken, C. (2014). A Framework Study for Healthcare Information System. Computer and Communications .

Challacombe, K. (2016). Health Research Data Repository (HRDR) overview.

Coeli, C. M. (2012). Achievements and Challenges for Employing Rrecord Llinkage Techniques in Health Research and Evaluation in Brazil. Epidemology application .

Coronel, P. R. (2009). Database Systems: Design, Implementation, and Management, Eighth Edition. Boston.

Cousins, J. (2002). What is Business Process Design and Why Should I Care?

Creamer, A. T. (2014). Research Data Management and the Health Sciences Librarian.

Dennis, A.(2012).System analysis and design.

dig_metast.pdf. (2007, Feburary 8). Retrieved October 13, 2017, from Dublin Core Metadata Guide Indiana Memory Project: https://www.in.gov/library/files/dig_metast.pdf

Dodero. (2008). A System Engineering Analysis Method for the Development of Reuse Computer-Supported Learning System.

Doiron, J. (2015). Health Research Data Repository (HRDR) overview.

Dong Joon Lee. (2017). Practices of Research Data Curation in Institutional Repositories: A qualitative view from repository staff. PLOS .

Eder, J. (2012). IT Solutions for Privacy Protection in Biobanking. In IT Solutions to Privacy Concerns.

EPHI. (2016). Ethiopian Public Health Institute. Retrieved January 16, 2017, from About EPHI: www.ephi.gov.et

Feldman, M. (2017). A Methodology for Quantifying the Effect of Missing Data on Decision Quality Cassifaction Problems.

FMOH. (2015). Health Sector Transformation Plan (HSTP). Addis Abeba, Ethiopia.

Franke T. (2013). Functional Requirements for a Central Research Imaging Data Repository.

Gavin, A. (2011). Key Issues for Consideration in The Development of a Data Strategy: A Review of the Literature.

Green, A. (2009, May). Policy-making for Research Data in Repositories: A Guide. UK.

Grozev, N. (2013). Performance Modelling and Simulation of Three-Tier Applications in Cloud and Multi-Cloud Enviroments. The Computer Journal, 2013 .

Hai Hua, M. C. (2011). A Data Warehouse for Translational Research.Journal of Biomedical Informatics .

Hank, C. (2016). Data Curation Profiling of Biocollections.

Hartson HR. (2003). Usablity Evaluation Methods.

Haux, R. (2006). Health Information Systems - Past, present, Future. International Journal Medical Informatics .

Health science. (2017, Aug 16). Retrieved from Data repository: http://guides.library.ucla.edu/health_data_statistics

Herbst,K.(2015).The INDEPTH Data Repository: An International Resource for Longitudinal Population and Health Data From Health and Demographic Surveillance Systems. Journal of Empirical Research on Human Research Ethics .

Hrynaszkiewicz, I. (2016). Publishing Descriptions of Non-public Clinical Datasets: Proposed Guidance for Researchers, Repositories, Editors and Funding Organisations.

Huang, R. (2013). Research of Improving the Quality of the Object-Oriented System. International Journal of Information and Education Technology.

IIBA. (2009). International Institute of Business Analysis: Non functional requirement.

Inmon, W. H. (2005). Building The Data Warehouse. New york: Wiley Computer Publishing.

ISO 9241-11. (2011). Ergonomics of Human-system Interaction Part 11: Usability: Definitions and Concepts.

Jackson, S. A. (1999). Data Warehouse Project White Paper.

Jacobson, I. (2006). Using UML Software Engineering with Object and Components, Second edition.

James J. Cimino, E. J. (2010). The Clinical Research Data Repository of the US National Institutes of Health . Laboratory for Informatics Development, National Institutes of Health Clinical Center, Bethesda, Maryland, USA .

Jana S. (2015). Rethinking Security of Web-Based System Applications.

Joseph M. (2013). Significance of Data Warehousing and Data Mining in Business Applications. International Journal of Soft Computing and Engineering (IJSCE) .

Justin Chisenga. (2006). The Development and Use of Digital Libraries, Institutional Digital Repositories and Open access Archives for Research and National Development in Africa: Opportunities and Challenges. WSIS Follow-up Conference on Access to Information and Knowledge for Development. Addis Ababa.

Koshak, N. (2002). Object-Oriented Data Modeling and Warehousing to Support Urban Design.

Kossi, E. K. (2010). Comparing Strategies to Integrate Health Information Systems Following a Data Warehouse Approach in Four Countries.

Kotari. (2004). Research Methodology Methods and Techniques,Jaipur.

Lee, D. J. (2017). Practices of Research Data Curation in Institutional Repositories: A Qualitative View from Repository Staff. PLOS ONE .

Little, G. (2003). Perceived Influences On Implementing Data Warehousing. In IEEE Transaction On Software Enginerring. http://dx.doi.org/10.1109/TSE.2003.1191794.

Meyer, B. (1988). Object-Orinted Software Constraction, Second Edition.

Naiburg. (2002). Literature Review and Classification of Electronic Commerc Research.Information Management.

Naiburg, E. (2001). UML for Database Design. Addison Wesley.

Nurmela, M. (2015). Aalto Data Repository Research Data Management, Sharing and Publishing in the World of Data Intensive Science. Abstract of master's thesis .

O'leary, z. (2004). The Essential Guide to Doing Research.

Oluwatosin, U. (2016). Usability and Security in User Interface Design: A Systematic Literature Review. Information Technology and Computer Science .

Omar. (2017, march 19). Object-Oriented Analysis And Design Introduction (Part 1). Retrieved from Medium.com: https://medium.com/omarelgabrys-blog/object-oriented-analysis-and-design-introduction-

Orr, K. (2000). Data Warehouse Technology: Revised Edition 2000, White Paper,.

Pamnani, R. (2015). Object Oriented UML Modeling for ATM Systems .

Pampel, H. (2013). Making Research Data Repositories Visible: The re3data.org Registry.

Pan, H. (2013). Sample and Data Sharing: Observations from a Central Data Repository.

Pandey, D. P. (2015). Research Methodology: Tools and Techniques.

Peer, L. (2012). Building an Open Data Repository for a Specialized Research Community: Process, Challenges and Lessons . The International Journal of Digital Curation .

Perkins, A. (2003). A Steratigic Approch Data Warehouse Engineering.

Peter Bugeia. (2013). 4A Data Management Acquiring, Acting-on, Archiving & Advertising research data at the University of Western Sydney. sydney.

Rahm, E. (2000). Data Cleaning: Problems and Current Approaches.

Rani, M. (2012). Systematic Archiving and Access to Health Research data: Rationale, Current Status and Way forward. Bulletin of the World Health Organization.

Rouse, M. (2005, April). Repository. Retrieved october 5, 2017, from http://searchoracle.techtarget.com/definition/repository

Rujirayanyong, T. (2006). A Project-Oriented Data Warehouse for Construction. Automation in Construction · November 2006 .

Scotte M. (2004). IBM Developer Works Requirments an Introduction.

Shelly, G. B. (2012). System Analysis and Design. Botson.

Silberschatz, A. (2010). Database System Concepts.

Sommerville, ©. (2004). Software Engineering, 9th edition. System models.

Tian J. (2006). Web Testing for Reliablity Improvement .

Tossy, T. (2014). Major Challanges and Constraint Integration of Health Information System in Africa Countries: A Nambia expriance.

Ucl Libraries. (2017, May 10). dr_sciences. Retrieved October wendnesday, 2017, from https://guides.lib.uci.edu/datamanagement

UN. (2003). Indicators for Monitoring the Millennium Development Goals. Definitions, Rationale, Concepts and Sources .

University of Manitoba. (2016, November 24). Manitoba Population Research Data Repository. Retrieved september 30, 2017, from viewDefinition.php: http://mchp-appserv.cpe.umanitoba.ca

Uzuegbu, C. P. (2012). Academic and Research Institutions Repository: a catalyst for access to development information in Africa . Helsinki.

Vardigan, M. (2015). Addressing Global Data Sharing Challenges. Empirical Research on Human Research Ethics .

Weichbroth, P. (2015). User Interface Prototyping. Techniques, Methods and Tools. User interface prototyping .

WHO. (2012). Systematic Archiving and access to health research data: rationale, current status and way forward. Manila.

WHO. (1990). The Role of Health Research.

Wordofa, T. T. (2011). Digitalization and Digital Preservation Activities and Initiatives at The Addis Ababa Univeristy: Towards Institutional Digitalization and Digital Preservation Strategy. Second International Conference on African Digital Libraries and Archives (ICADLA-2) .

Zhang, Y. (2015). Data Management and Curation Practices: The Case of using DSpace and Implications.

Annexes

Annex 1

ADDIS ABABA UNIVERSITY
SCHOOL OF PUBLIC HEALTH AND SCHOOL OF INFORMATION SCIENCE
DEPARTMENT OF HEALTH INFORMATICS

Information Sheet

My name is _____. I am a student at Addis Ababa University and I am currently conducting a project for the partial fulfillment of master's degree in health informatics. The purpose of this project is designing web-base integrated health research data repository Ethiopian Public Health Institution.

I would like to ask you some questions related to the topic I mentioned above and all of your responses to questions will be kept strictly confidential throughout the project. Participation in this project is voluntary and you can choose not to answer any individual question or all of the questions. I look forward for your full participation as the answers you give on this form will help in better understanding of the situation of the research data handling of the institution and will help to designing appropriate technology to alleviate the existing system problems regarding health data. Your participation, non participation or refusal to answer the questions will have no effect now or in the near future on your professional activities and personal life.

May I get your permission to continue?

Yes ◯ Go to the consent form

No ◯ Stop -

Interview questions guiding

Name:_____ Gender:_____ Profession:_____

1. Description of the data

1. Approximately, how many data files exist?

---

2. What is the average size of the data files? (units) or overall (total file size)

---

3. What format(s) are the data stored in?

---

2. Data Flow & Use

1. How was the data acquired/collected?

---

2. What specific software programs or tools/hardware were used in the collection/generation of the data?

---

3. What specific software programs or tools/hardware are required to utilize this data?

_____

4. Describe briefly the way the data is currently organized (i.e., file name conventions, existing metadata, units)?

---

3. Storage

1. Where are the files currently stored? Include the storage media(s) and any tools used in your management of the data.

---

2. Are there backups of the data?

---

3. Who is primarily responsible for managing these files?

---

4. Stakeholders/researchers/ officials

1. Who is the intended audience of this data? Is the data intended to be made available for re-use by others?

---

2. Who might you imagine would be interested in this data? (e.g., other researchers in my field, researchers outside of my field, practicing professionals, policy makers, etc.)

---

3. How might this data be used by these people?

---

5. Jobs

1. What is your current job title?

---

2. How many years in total have you been working in your current job

---

3. Describe your work setting?

---

4. Please indicate your credentials and degrees.

---

5. Please provide any other educational or training you have received that is applicable to performing your job.

Annex .2

User interface evaluation questionnaire, Adopted from John Brooke john.brook@redhatch.co.uk

Use X mark for your answer

No	Questioners	Strongly disagree	Disagree	Neutral	Agree	Strongly agree
1.	It was easily manageable options of the system					
2.	It was clear to access the different options					
3.	The system provided adequate functions					
4.	It was clear on using the different functions of the system					
5.	The system flexible in allowing the user to choose options					
6.	The system handle errors in user actions					
7.	Overall, the interface was user friendly and attractive					