IJCSE International Journal of Computer Sciences and Engineering Open Access
Research Paper Vol. 10, Issue.6, June 2022 E-ISSN: 2347-2693

Study on Theoretical Aspects of ontology-based and Virtual Data Integration in medical intelligence process and its Applications

Asogwa E.C.[1], Amanze B.C.[2*], Ngene C.C.[3], Belonwu T.S.[4], Chukwuogo O.E.[5]

[1,3,4,5]Department of Computer Science, Faculty of Physical Sciences, Nnmadi Azikiwe University, Awka, Nigeria
[2]Department of Computer Science, Faculty of Physical Sciences, Imo State University, Owerri, Nigeria

**Corresponding Author: amanzebethran@yahoo.com, Tel: +2347062523438*

DOI: https://doi.org/10.26438/ijcse/v10i6.3745 | Available online at: www.ijcseonline.org

Received: 23/May/2022, Accepted: 12/Jun/2022, Published: 30/Jun/2022

Abstract- Lack of fast, accurate, reliable and intelligent software solutions that can help healthcare practitioners make decisions that would solve urgent, and in some cases, complex medical problems in real-time. Cost of processing and analyzing large volumes of data in a medical environment is high most especially in terms of time consumption. Application model for enhanced medical intelligence process were developed in this paper and it can be applied in healthcare centers, clinics and maternities in Nigeria. The healthcare centers, clinics and maternities, etc can link the application model developed to their database servers so that the application will connect the platform to the database server in other to carry out disease control procedures using ontology-based (OBDI) and virtual data integration (VDI) techniques as the have the ability to ensure abstraction of data that comes from multiple sources in varying schemas, syntactic accuracy and to have a seamless transition from data into information, then into action. The objective of the design is to develop an application model for enhanced business intelligence process which was achieved using ontology-based data integration (OBDI) system, application model uses intelligent agent to guide doctors accurately by carrying out disease control procedures. Test results on the new system using confusion matrix shows a significant positive impact 88% accuracy in medical intelligence process as against 60% of accuracy by the existing system, and hence a significant improvement on overall operating efficiency. The model is therefore recommended for use by Physicians, government, hospital administrators and patients.

Keywords- OBDI, Hospital administrators, database, VDI, DV and Physicians

I. INTRODUCTION

Data integration is the process of integrating data from multiple sources and probably has a single view over all these sources. It also involves the answering of queries using the combined information. It can be physical-coping the data to warehouse or virtual-keep the data only at the sources. Heterogeneity is integration's major challenge as data can be stored with different systems such as relational database, web database, xml database, etc. The problem of data integration was and still remains the extremely difficult part of data management field.

The process of extracting data from production data stores, cleansing them, transforming them, correlating them is generally known as data integration. In data virtualization the data is not moved physically as part of data integration, but all the data cleansing, data transformation and data correlation is defined in logical layer that is then applied to data as they are fetched from original data source while generating reports. It has better agility and shorter implementation cycle thus enabling more analysis and hypothesis testing. It attempts to perform data cleansing, data transformation and data correlation as data moves out from production systems thus avoiding any intermediate storage. As oppose to data warehouse approach that physically changes data in each stage and loads it into some data store. The cleansing, transformation and correlation are defined programmatically using SQL like query language [1].

Application of data integration implies we consider whether data needs to be physically moved or whether a virtual or "in-place" approach to accessing and aggregating data makes more sense. This virtual data integration is the approach this research is applying in combination with ontology-based technique. Solid data integration strategy can help us make more effective decisions based on trusted, complete data and transform the way our business works. The need for systematic data integration is more eminent since the amount and heterogeneity of data is not declining but increasing constantly. A data integration scenario starts with identifying the data sources that will participate in the applications and then build a virtual schema (often called a mediated schema), that would be queried by users or applications. Querying process would begin by reformulating a query posed over the virtual schema into queries over the data sources and then executing it efficiently with an engine that created plans that span multiple data sources and deal with the limitation and capabilities of each source. The need for real-time analytics and reporting resulting from competitive and compliance pressures are driving the need for a virtual data warehouse approach to integrate and present disparate data. There are three types of data integration methods; data consolidation, data propagation and data federation sometimes referred to as virtual data integration (Data virtualization). The data federation provides a single virtual view for two or more data sources. The virtual view creation among different data sources is still a challenging task due to their heterogeneities. Data Virtualization (DV) is any approach to data management that allows an application to retrieve and manipulate data without requiring technical details about the data, such as how it is formatted or where it is physically located. Unlike the traditional extract, transform, load ("ETL") process, the data remains in place, and real-time access is given to the source system for the data, thus reducing the risk of data errors and reducing the workload of moving data around that may never be used. Unlike Data Federation it does not attempt to impose a single data model on the data (heterogeneous data). The technology also supports the writing of transaction data updates back to the source systems.

To resolve differences in source and consumer formats and semantics, various abstraction and transformation techniques are used. This concept and software is a subset of data integration and is commonly used within business intelligence, service-oriented architecture data services, cloud computing, enterprise search, and master data management. Data Virtualization software is an enabling technology which provides some or all of the following capabilities:

- Abstraction – Abstract the technical aspects of stored data, such as location, storage structure, Application Processing Interface (API), access language, and storage technology.
- Virtualized Data Access – Connect to different data sources and make them accessible from a common logical data access point.
- Transformation – Transform, improve quality, reformat, etc. source data for consumer use.
- Data Federation – Combine result sets from across multiple source systems.
- Data Delivery – Publish result sets as views and/or data services executed by client application or users when requested.

The software may include functions for development, operation, and/or management.
Data federation is an aspect of data virtualization that employs a variety of specialized query techniques to draw the right information from operational systems quickly for immediate analysis without negatively impacting the performance of source systems.

[2], there is a slight difference in the definition and explanation of data virtualization and data federation. Data virtualization (virtual data integration) which is the technique this research is applying, is defined as the process of offering data consumers a data access interface that hides the technical aspects of stored data, such as location, storage structure, API (application processing interface), access language, and storage technology. It provides an abstraction layer that data consumers can use to access data in a consistent manner, which hides all the technical aspects of data storage. The applications don’t have to know where all the data has been stored physically, where the database servers run, what the source API and database language is and so on. While the same author defined data federation as a form of data virtualization where the data stored in a heterogeneous set of autonomous data stores is made accessible to data consumers as one integrated data store by using on-demand data integration.

In terms of consistency and data integration, in the virtual approach to data integration, one usually assumes that certain integration consistencies (ICs) hold at the global level. However, there is no global integration consistency (IC) maintenance mechanism. Actually, under the local as view (LAV) approach they are sometimes used to generate a query plan to answer a global query. There are situations where without assuming and using those global integration consistencies (ICs) no query plans can be generated [3]. Furthermore, the virtual data integration approach leaves the information requested in the local sources. It will always return a fresh answer to the query. The query posted to the global schema is reformulated into the formats of the local information system. The information retrieved needs to be combined to answer the query. Technically, data virtualization can be implemented in many different ways. Some of them include:

- With a federation (data virtualization) server, multiple data stores can be made to look as one. The applications will see one large data store, while in fact the data is stored in several data stores.
- An enterprise service bus (ESB) can be used to develop a layer of services that allow access to data. The applications invoking those services will not know where the data is stored, what the original source interface is, how the data is stored, and what its storage structure is. They will only see, for example, a SOAP or REST interface. In this case, the ESB is the abstraction layer.
- Placing data stores in the cloud is also a form of data virtualization. To access a data store, the applications will see the cloud API, but they have has no idea where the data itself resides. Whether the data is stored and managed locally or whether it’s stored and managed remotely is transparent.
- In a way, building up a virtual database in memory with data loaded from data stored in physical databases can also be regarded as data virtualization. The storage structure, API, and location of the real data is transparent to the application accessing the in-memory database. In the business intelligence (BI) industry, this is now referred to as in-memory analytics.
- Object-Relational Mappers (ORM) are tools that convert data structures from data stores to the concepts

used in an object-oriented programming model, such as Java and C#. Example an ORM might be able to convert the flat table structures of an SQL database to the Object-Oriented concepts used in Java. The effect is that the Java program doesn't have to understand and deal with the characteristics of the SQL concepts but purely with Java concepts. Examples of ORMs are Hibernate, NHibernate and iBATis.

- Organizations could also develop their own software-based abstraction layer that hides where and how the data is stored.

This research intends to implement the first so as to achieve the data virtualization (virtual data integration) phase since our focus is on medical intelligence systems. Data virtualization servers (DVS) is a dedicated product designed to support data virtualization, meaning it can present multiple heterogeneous data stores as one logical data store to data consumers. Its accessing process is similar to logging on to a database server. Also, Data virtualization servers generally have their own internally technical architecture, but on a conceptual level, all of them consist of at least two main modules, the design module and the runtime module. The design module is used by analyst, designers and possibly users to enter virtualization specifications. This is where specification such concepts definitions, data models, and specifications for transformation, cleansing and integration are entered. All the above is stored in a dictionary. While the run-time module is the module used when data consumers access the virtualization layer, it's the module that handles all the in-coming requests. It determines the best strategy for running the queries, knows how to access different data store technologies, and knows how to integrate data etc. It consists of a scheduler, a storage engine, cache manager, a query optimizer and so on. In addition it also has a dictionary for storing meta-data specification.

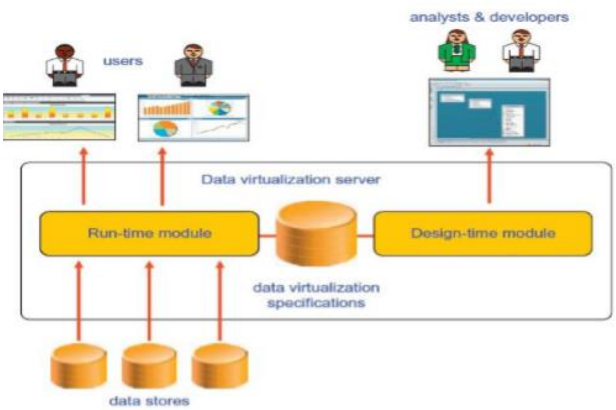

Figure 1: A Data Virtualization server is made up of a design module, a runtime module, and a dictionary for storing all the specifications. [2]

Virtualization

Virtualization helps in dealing with data intensive problems. The virtualization process is defined as the mapping of an abstract data set to a virtual space according to three major intertwined steps consisting of data selection for representing the problem space, assumptions definition to define the final virtual space, the mapping between the starting space and the final space through a metaphor [4]. Virtualization features in the management of distributed data with data intensive problems such as;

- The preservation of data and knowledge in their actual format. The virtual process has no impact on the physical reality. Its putative evolution; that is data and knowledge production can go their own way without any necessary change.
- The operation-ability implies virtualization is not just abstraction, it allows to recursively transform the physical reality according to the lessons learned in virtual reality. That is it aims to actualize physical reality, such that any change in physical reality has its counterpart in virtual reality.
- As par heterogeneity especially on the internet, virtualization was a successful solution to the problem. Because protocol heterogeneity of existing computer communication architectures was not able to interoperate although they were built for the same purpose that is for exchanging data between distant computers. Virtualization process allows to integrate existing communication technologies and to preserve the future development.
- Portability- Virtualization process allows for portability as it allows software to be run on any microprocessor architecture without re-writing it, but just compiling it with ad-hoc compilation.
- Virtualization can be mathematically represented formally by sets say E and V corresponding respectively to physical reality and virtual reality and the two functions ME and MV.

$ME: E \rightarrow V,\ MV : V \rightarrow E$ …………………………..(1)

ME and MV define the mapping rules between E and V. V is the metaphor that refers to a domain of knowledge [4]
Virtualization process has being successfully used by information technologies to deal with heterogeneity problems, to increase productivity of information technology tools and spread information technology products to non-information technology skilled users.

Data Virtualization layer isolates Business Intelligence tools from the details for accessing the underlying data sources. Through the data virtualization layer a Business Intelligence tool can get access to any data available either in the data warehouse, staging area or in the production systems, enabling for instance single views of customer with information about a customer coming from the customer relation management (CRM) system, historical information about this customer retrieved from the data warehouse, billing information from the billing system, etc.

a) Data Virtualization enables access to real-time data.
b) Data Virtualization enables shorter time to market for delivering reports to users.
c) Data Virtualization enables access to any data type from semi-structured to unstructured sources, internal or external, including recently arising sources such as non-SQL databases.

There is cache in data virtualization layer that can be used to store any data to keep track of historical information when needed [5]

Data Integration System Model

One of the most important aspects in the design of a data integration system is the specification of the correspondence between the data sources and those in the global schema. Such a correspondence is modeled through the notion of mapping as introduced in the previous section. It is exactly this correspondence that will determine how the queries posed to the system are answered. Two basic approaches for specifying the mapping in a data integration system as stated in previous section are called local-as-view (LAV), and global-as-view (GAV). Assuming S1,..., Sn to be the local schemas of n pre-existing data sources and G1,...,Gm to be m global relations of the global schema G. The aim is to model semantic relations between the local schema Si and the global schema Gj [6]

In Local-as-view (LAV) as shown in figure 2, the semantic mappings are of the form

$$Si \subseteq Vi\ (G1, \ldots, Gm)$$

Where each Vi is a view over the global schema, i.e., a query built on global relations. In this approach, the source schemas are modelled as a set of views over an underlying global schema. The local-as-view method, does not describe the global schema in a direct manner but in a way which is disassociated from the data sources. The mapping between the global schema and the data sources is created by creating views of each source over the global schema. The advantage of this model is that new sources can be added easily when compared to GAV.

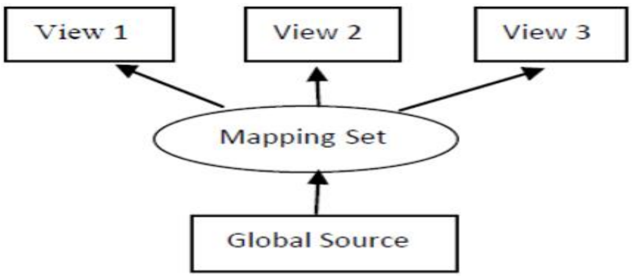

Figure 2: Local-As-View (LAV) [6]

While in Global-as-view (GAV) as shown in figure 2, the semantic mappings are of the form

$$Vi(S1,\ldots,Sn) \subseteq Gi \qquad (2)$$

Also equivalently denoted as

$$Gi \supseteq Vi(S1\ldots,Sn) \qquad (3)$$

Where each Vi is a view over the local schemas, i.e. a query built on local relations. In this approach, the global schema is modeled as a set of views over the source database. The advantage of this model is that the query processing is simpler. But the disadvantage is that the addition of a new source needs considerable effort. This makes GAV a good choice when the sources are less probable to change.

2. Statement of the Problem

In healthcare system all over the world, the amount of patient oriented data is constantly growing. More hospitals are opening up with various departments / units. For example, the Intensive Care Unit (ICU) is an extremely data intensive environment where large volumes of data from patient monitoring and observations are recorded continuously. Such patient oriented data could be generated from medical devices, laboratory results, electronic prescriptions, therapeutic decisions, and clinical observed values by physicians and nurses. These data is disintegrated and access to the data has to been done by requesting for it through the various hospital units. This is not only time consuming but obsolete. The following are the existing challenges of the medical system;

1. Lack of fast, accurate, reliable and intelligent software solutions that can help healthcare practitioners make decisions that would solve urgent, and in some cases, complex medical problems in real-time.
2. Cost of processing and analyzing large volumes of data in a medical environment is high most especially in terms of time consumption.

So this thesis proposed a patient oriented design for integration of large volumes of data in order to improve database validity compared to procedure oriented design that multiplies the redundancy of data. The paper is to address the problems, by having a hybrid of ontologies and virtual data integration in order to enhance medical intelligence process.

3. Aim and Objectives of the Study

The aim of this paper is to develop an application model for enhanced medical intelligence process.

The objectives include:

i. To design a database system for storing medical records.

ii. To develop software for enhanced Medical Intelligence Process that would be more user-friendly, flexible, adaptive, intelligent, agile and automatic in integrating and analyzing medical data thereby helping medical practitioners at various levels to make realistic intelligent and real-time decision on critical health issues.

iii. To develop a system that will integrate the various patients medical records from different hospitals using ontology based and virtual data integration technique that will allow clinic data of one patient collected together to form a combinational resource, and could be accessed by physician if authority is assigned to the physician

4. Significance of the Study

Physicians, government, hospital administrators and patients will benefit immensely from the hybrid model for enhance medical intelligence.

i. The physicians will have an intelligence system that will help them in their medical practice and also in diagnosing and treating patient's ailment more quickly and effectively.

ii. The hospital administrator can process information on real-time based and this will remove redundancy and delay in retrieving data.
iii. Patient's healthcare conditions will be improved as they will access quicker and more accurate diagnosis and treatment.
iv. Government will benefit from the new system as healthcare system will be greatly improved leading to a healthy and wealthy nation.

Ontology-Based Data Integration Methods

Data integration methods generally depends on which integration architecture the developer is quite familiar with, how much is known about heterogeneity of data sources, etc. In the models shown in figure 3 selected criteria for analyzing is used, ontologies and data sources are represented as classes and semantic relationships between data source content and ontology as links and links between local ontologies as mappings.

The data integration task automatically performs data extraction and integration from both structural and semi-structured data sources. The proposed model integrates and reuses data using ontologies by relevant criteria. As earlier stated, ontology is a data model which consists of classes, properties and relationships between them as its parts. Its power lies in the ability to represent relationships between the classes. The benefits of the model mainly are its runtime interpretation, an open-world assumption and its ability to clearly interpret disseminating of knowledge between people and applications. It supports integration task as it describes the exact content and semantics of data sources more explicitly. Also, in contrast to database schemas which are static, ontology schemas are highly dynamic and are an evolving object [7]

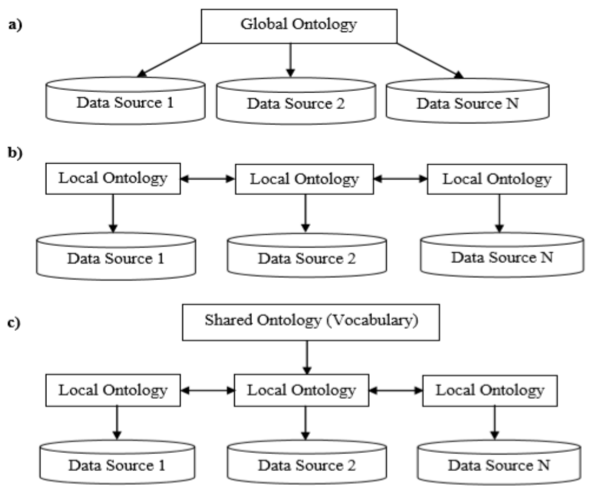

Figure 3: The Methods: a) Global, b) Multiple c) Hybrid [7]

The global ontology as shown in figure 3a) uses a single ontology which as a single main stage of building global ontology by domain expert that knows the semantics for all data sources. It can also be a combination of several specialized ontologies. It describes data from heterogeneous data sources from which query is executed via the main ontology. The every data source in this method is related. It is applied to integration solutions where all data sources to be integrated provide the same view on a domain.

The b) of figure 3 is the multiple ontology method which uses local ontologies and mapping rules between them. Each data source is described by its own ontology and the mapping rules can be modified according to the dynamic change of data source. It has two main stages; building local ontologies and defining mappings. This method local ontology describes data from heterogeneous data sources such that integrated query is executed via the local ontologies. The essential feature of this method is that the ontologies for individual data source could be developed or changed without respect to other semantic relations, data sources or their ontologies.

In the case of the hybrid ontology method c) of figure 3, it uses a vocabulary of a domain to represent a shared ontology, a local ontology and mapping rules between them. The specification of a vocabulary includes definitions of classes, relations, functions, and other objects. The mapping rules can be modified according to the dynamic change in data source. That is, it integrates users' profiles, data warehouse, online Analytical Processing, Data Mining and underlying Enterprise Integration System (EIS). It aims to develop mechanisms for smooth mapping from user-defined keywords to meta-data items in data warehouse or physical attributes and entities dispersing in business system service or office system service (BSS/OSS). This architecture requires ontological engineering techniques for effective operation, such as building ontological namespace and semantic relationships for organizing items in the domains; mechanisms for ontological transformation and mapping intra or inter domains; services for ontological query and search in the warehousing [8]. The advantage of this hybrid method which this research will implement is that new data sources can easily be added without the need to modify the mappings or the shared vocabulary. So this architecture gives more autonomy to data sources. The use of shared vocabulary makes the source ontologies comparable and avoids the disadvantages of single or multiple ontology methods. Table 1 gives a summary of the benefits and drawbacks of the ontology-based integration methods.

Table1: Advantages and Disadvantages of ontology-based integration methods [7]

Criterions	Ontology-based Architectures		
	Global	**Multiple**	**Hybrid**
Evaluation of Semantic Heterogeneity	Useful for Systems which have the same view on a domain	Useful for Systems which have the same on a domain	Useful for Systems which have different view on a domain.
Appending new data	Some modificatio	Supports an opportunity to	New data sources can

sources	n is necessary in the global ontology	append the new data source with some adaption in other ontologies	easily be added without the need of modification.
Elimination of data sources	Some modification is necessary in the global ontology	Supports an opportunity to remove the data source with some adaption in other ontologies (need to remove relation between ontologies) Data	Data sources can easily be removed without the need of modification
Comparison of multiple ontologies	Impossible	Difficult because of lack of a common vocabulary	Simple because ontologies use a global shared vocabulary

Ontology-based data integration is an effective method to cope with the heterogeneous data. This solution is based on the idea of decoupling information semantics from the data sources. Moreover, ontologies support dynamic domains better. For this reason, it is necessary to analyze data source elements: data, schema, schema elements and content, values, entities and attributes, query result classes. It is known that ontology-based search system gives the user more meaningful query results than the normal search system, which queries data with syntactic parameters. The query result is based on data retrieval methods [7]

Ontology Integration is similar to Ontology merging only that in ontology integration the integrated ontology is created reusing parts of source ontologies as they are. Both have a key task of in consistency checking process which must ensure the absence of unforeseen or wrong implications into the integrated ontology.

The ontology is created by using Resource Description Framework (RDF), Resource Description Framework Schema (RDFS), DARPA Mark-up Language (DAML) + Ontology Interchange Layer (OIL) and Ontology Web language (OWL). Among these, the OWL is more powerful than others. The OWL has well defined semantics and highly optimized implementation system. The data quality is often defined as “fitness for use”. Data is fit for use whenever a user, (1) is able to get information, (2) is able to understand it, (3) finds it applicable to a specific domain and purpose of interest and (4) believes it to be credible. The Key measures of data quality are data completeness, data consistency and data accuracy. Completeness is defined as the extent to which data are of sufficient granularity for the task at hand. Data consistency expresses the degree to which a set of data satisfies a set of integrity constraints. Data accuracy is defined as the closeness between the given value and the correct representation of the same in real life phenomenon.

A unified view is created to resolve the semantic conflict among different heterogeneous databases by using ontology. This view is used by the user for shopping and business analysts for decision support [9]

This ontology technique will provide quick access to documents and information with the help of taxonomy created from the concepts, called a concept map, with the incorporation of ontology array indexing. The system will be adaptive as it will retrieve the most relevant information as well as documents which are close to the user’s queries and the array-indexing is key to this concept as it helps in obtaining an inter-relation between the documents and information. It also, ensures the redundancy of concepts or document in the database.

Business intelligence requires the acquisition and aggregation of key pieces of knowledge from multiple sources in order to provide valuable information to customers or feed statistical business intelligence models and tools. Since the analyst of business intelligence has lots (massive) of information to extract and process, there is the challenge as par its acquisition and use of its semantic information, ontology based data integration approach is applied in this research as an effective way to resolve the challenge. It implies the process of identifying in text or other sources relevant concepts, properties, and relations expressed in ontology [10].

As an intelligent system like humans which implies having the ability to take in data from their environment, understand the meaning and significance of the information and then act appropriately, business intelligent and analytics has to do with integrating all information streams produced by a firm into a single, coherent enterprise-wide set of data, and then, using modeling, statistical analysis tools (normal distribution, correlation and regression analysis, chi square analysis, forecasting and cluster analysis), and data mining tools (pattern discovery and machine learning), to make sense out of all the massive data so managers can make better decisions and better plans, or at least know quickly when their firms are failing to meet planned targets.

The tools that use ontologies for data integration can be compared on the following ways;
1. The use of ontologies which implies the different roles ontologies play in the data integration process in terms of making the context explicit and in the expression semantics or as a global query model.
2. Ontology representation which has to do with representational capabilities of the used ontologies that can vary amongst the different tools.
3. Use of Mapping has to do with linking the ontologies to the actual information and to each other if multiple ontologies are used.
4. Ontology engineering that deals with the building and re-use of ontologies in different manners.

II. REVIEW OF RELATED WORK

[11] Presented Ontology-based Clinical Reminder System to Support Chronic Disease Healthcare. He said that improving quality of healthcare for people with chronic conditions requires informed and knowledgeable healthcare providers and patients. Decision support and clinical information system are two of the main components to support improving chronic care. One describes an ontology-based information and knowledge management framework that is important for chronic disease care management. Ontology-based knowledge acquisition and modeling based on knowledge engineering approach provides an effective mechanism in capturing expert opinion in form of clinical practice guidelines. The framework focuses on building of healthcare ontology and clinical reminder system that link clinical guideline knowledge with patient registries to support evidenced-based healthcare. He describes implementation and approaches in integrating clinical reminder services to existing healthcare provider environment by focusing on augmenting decision making and improving quality of patient care services. The paper was focused on clinical reminder service and didn't integrate electronic health record (EHR) standards and this is the research gap established in his work. [12] titled a data integration platform for patient-centered e-healthcare and clinical decision support", it proposed an open data integration platform for patient, clinical, medical and historical data across multiple health information systems. As an open platform, it can accommodate and integrate further heterogeneous data sources such as data streams generated by wearable Internet of Things (IoT) devices. As an integration platform, it facilitates centralization of data assets. This centralization empowers every stakeholder in a patient-centered care setting to actively participate in decision-making. A range of analytics and reporting solutions, such as data warehouse, interactive dashboards, and predictive analytics tools, can be deployed upon this open data integration platform. The proposed platform is currently being adapted and implemented to address patient-centered healthcare and clinical decision support requirements in a sports injury clinic at a not-for-profit private hospital in Melbourne, Australia. Use Case based demonstration of the platform's suitability for holistic information management, decision-support, and predictive analytics justify its role in the advancement of e-healthcare. The research work suggested than an advanced analytics, data visualization, monitoring and reporting functionalities for clinical decision support will should be added and customized in future work and this is identified as the research gap. Medical decision support systems based on machine learning was presented by [13]. The central idea of the dissertation is aimed at facilitating personal health care, reducing costs of health care, and improving outcomes. They proposed a new machine-learning algorithm for three disjointed health care problems: hospital referral, cost-effective diagnosis, and lifestyle recommendation. In the hospital referral and lifestyle recommendation projects, individualized recommendation is generated based on the input of personal characteristics and preferences. The systems can then return the best individual solution (hospital selection or the plan of lifestyle changes) that fits one's preference and personal considerations. In the cost-effective diagnosis project, the recommendation of a test is provided based on individual information (including symptoms and previous test results). The recommended test has the highest potential to cross (or get close to) the treatment (or non-treatment) threshold. In other words, they optimize diagnosis in terms of the number of tests and the amount of cost without sacrificing accuracy (sometimes improving accuracy).

III. METHODOLOGY

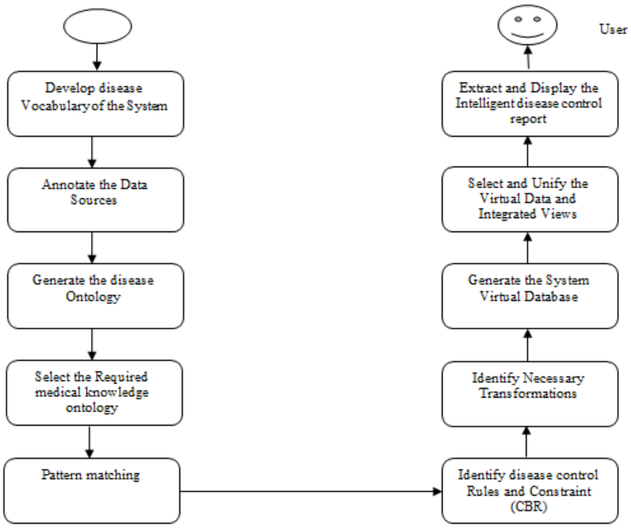

Figure 4: Steps of the Methods Used.

The method as shown in figure 4 entails the enhanced of the ontology-based and virtual data integration process. The process flow of the system would be boosted with the software agents as indicated in the data integration layer of the system design. The data integration layer is where the ontology-based and virtual data integration process takes place and it is the aspect of the business intelligence process that the research is enhancing.

Experimental Output Develop

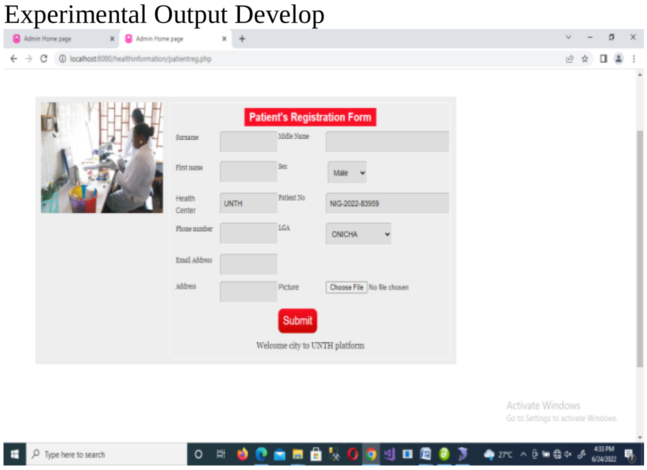

Figure 5: Patient Registration Form

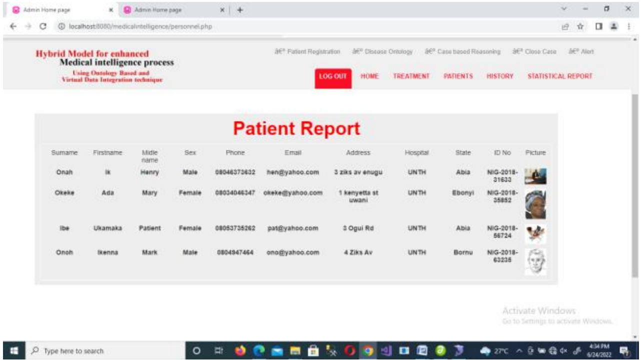

Figure 6: List of Patients

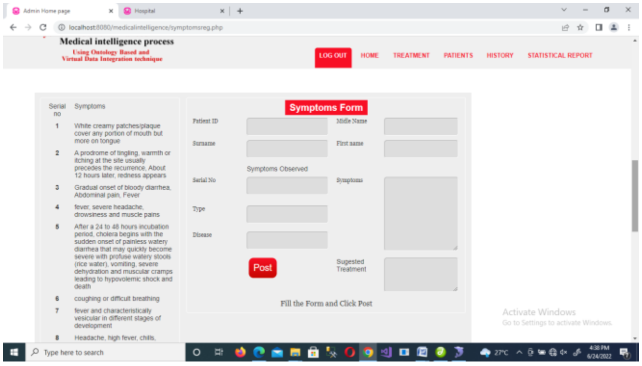

Figure 7: Disease ontology Form

IV. RESULT AND DISCUSSION

Confusion matrix applied to test dataset of Disease Control using Ontology-based data integration technique

Table 2: Confusion Matrix Observed

		True	**False**
Predicted	**True**	13	2
	False	5	10

Table 2 shows that out of 20 disease control test carried out using ontology-based data integration technique for disease control procedure; 13 diagnoses are True Positive and was predicted correctly, 2 diagnoses were true while they are not, 3 diagnoses was detected to be False while it is not thereby allowing the disease to go undetected. Finally, 2 False was detected where no disease was found and it was correct. A model of performance metrics can be derived from the confusion matrix as show in equation 4, which show the level of accuracy of the ontology-based data integration technique for disease control procedure.

Substituting the values we have

$$AC = (13+10) / (13+2+5+10) \quad \ldots\ldots(4)$$

AC = 0.88 i.e. 88% accuracy in predicting the disease control procedure using ontology-based data integration technique

V. CONCLUSION

In this paper, application model for enhanced medical intelligence process was developed. The business model developed focused on expert system for health sector that uses intelligent agent to guide doctors in accurately carrying out disease control procedures. The key concept of the developed business intelligence model adopted ontology-based (OBDI) and virtual data integration (VDI) techniques that the ability to ensure abstraction of data that comes from multiple sources in varying schemas, syntactic accuracy and to have a seamless transition from data into information, then into action. The model leverage the benefits of ontology-based and virtual data integration techniques in a business intelligence environment in order to provide intelligent, agile, adaptive, user-friendly, etc real-time solutions as well as structural (schema), syntactic (format) and semantic (meaning) heterogeneity correct data model to decision-making for business users. At the end of this paper, the system developed was used to manage a disease registry that consists of the concepts of the domain, the attributes characterizing each disease, the different symptoms, and treatments. Also a relational database for storing and tracking disease outbreak and control using ontology-based data integration (OBDI) was achieved. A platform for Virtualized Data Access – Connect to different data sources and make them accessible from a common logical data access point was created which integrated an intelligent agent that uses case based reasoning for determining the disease control procedure to be applied to patient treatment for effective control of the disease. For efficiency in disease control management, a statistical report of disease treatment records according to states or the country as a whole was generated in the system developed. From the system test conducted, we saw that the Ontology-based data integration technique for disease control procedure has 88% accuracy in predicting the disease control procedure.

REFERENCES

[1].Amineh, A., Hadi, S., Nasser, N. (2008). A RDF-based Data Integration Framework. NEEC 2008 www.1211.6273.pdf/ retrieved on **May 23, 2021**

[2].Rick, V. D. L., (2012). Data Virtualization for Business Intelligence Systems", www.r20.nl Retrieved from www.3-s2.0-B978...000010.pdf/ on **Dec.7, 2012**

[3].Leopoldo, B., (2007). Virtual Data Integration" Carleton University School of Computer Science Ottawa, Canada, www.tutorial-Bertossi.pdf/ retrieved on **May 5, 2021**

[4].Magali, R. & Michel, S., (2015) .Virtualization in System Biology: Meta Model & Modeling Language for Semantic Data Integration" retrieved on **May 29, 2021**

[5]. Francesco, D. T., Ezio, L. & Filippo, T., (2015) .Academic Data Warehouse Design Using Hybrid Methodology. Computer Science & Information System 12(1):135-160 DOI: 10.2298/c815140325087D www.csisn3p135-160.pdf/ retrieved on **May 5, 2021**

[6]. Munmun, B. & Nashreen, N. (2016). Study on Theoretical Aspects of Virtual Data Integration and its Applications. International Journal of Engineering Research and Applications, **6(2), 69-74, 2015.**

[7].Virginija, U. & Rimantas, B. (2011) . Ontology-based Foundations for Data Integration. The First International Conference on Business Intelligence and Technology Copyright IARIA, 2011. ISBN: 978-1-61208-160-1.

[8].Longbing, C., Chengqi, Z., & Jiming, L., (2017). Ontology-Based Integration of Business Intelligence. Retrieved from www.w639.pdf/ on **Dec.5, 2017.**

[9]. Hema, M. S. & Chandramathi, S., (2013). Quality Aware Service Oriented Ontology Based Data Integration. WSEAS Transactions on Computers E-ISSN: 2224-2872 12 (12), 12-16

[10].Mezghani, E., Exposito, E. , Drira, K., Silveira, M. and Pruski, C. (2015). A Semantic Big Data Platform for Integrating Heterogeneous Wearable Data in Healthcare," Journal Medical System, **39(12), 185, 2015**.

[11].Marut, B. (2016). Ontology-based Clinical Reminder System to Support Chronic Disease Healthcare. Article in IEICE Transactions on Information and Systems · DOI: 10.1587/transinf.E94.D.432 · Source: DBLP

[12].Madhura, J., Dinithi, N., Daswin, S., Damminda, A., Brian, D., Kate, E .W. (2020). A data integration platform for patient-centered e-healthcare and clinical decision support. Research Center for Data Analytics and Cognition, La Trobe University, Victoria, Australia b School of Allied Health, La Trobe University. Victoria, Australia

[13].Chih-Lin, C. (2019). Medical decision support systems based on machine learning. PhD (Doctor of Philosophy) thesis, University of Iowa, https://doi.org/10.17077/etd.o5gmwvxk

