CLINICAL INTELLIGENCE FRAMEWORK FOR DECISION-SUPPORT

By

FADI LOUIS NAMMOUR

A thesis submitted to the Department of School of Computing and Information Systems

In conformity with the requirements for

the degree of PhD

Kingston University

Kingston upon Thames, London

(12, 2018)

Abstract

The healthcare setting has been evolving in many ways to keep up with the changing technology and information challenges. The modern healthcare setting hosts patient and clinical data over many co-existing information systems and medical devices. It is rarely the case that these data sources correctly exchange clinical data in a standard way, and for that reason answering questions that come up during the management of the healthcare processes requires looking up the pieces of information in this distributed and loosely connected ecosystem.

As the healthcare organization grows in both size and operational activities, decision-making stakeholders become more dependent on status summaries, which become harder to manually prepare with large volumes of scattered data. As the decision-makers follow a performance-driven evaluation approach, their enquiries span multiple data sources and require adaptation of data models in many information systems and data silos. Decision-support officers and clinicians find themselves in need to work with technical personnel to help them connect to the many data systems and combine their data in a way the delivers the correct expected results to the stakeholders. As the demands are not the same every time, this process depends on continuous and expensive technical involvement.

The aim of this thesis is to enable non-technical clinicians and decision-support officers to work with data from different information systems and medical devices using a framework software system that consolidates this data and allows them to create their own analytical enquiries without the continuous assistance of technical personnel. To achieve this aim, the thesis critically examined current approaches and existing solutions, then identified the need for a new approach to eliminate the limitations identified, analysed these approaches and proposed a new approach. The proposed solution was designed and developed following well-established and tested methodologies and was then evaluated and its impacts were identified in the healthcare environment.

The research in this thesis questioned the possibility of having a consistently usable framework system where non-technical users can execute their technical analytics against clinical data from different data

sources in the healthcare setting. A case study was performed in multiple hospitals to collect the needs of different users and to identify the gap in the current situation. The results showed that the adoption process was possible and was tested through an application scenario where the users contributed collectively to the creation of a monthly status report portal. The portal was used by the stakeholders to follow-up on the healthcare setting's current and historical performances and they added new requirements in the form of aggregation requests that were executed through the analytical framework system.

The framework system was developed using a Model-Viewer-Controller software design approach and followed the best-practices of modern software engineering. The service-oriented architecture was adopted to govern the delivery of data from an application server to thin clients such as a web-browser or a mobile device. An algorithm was developed to enable the dynamic execution of user analytical enquiries against a dynamic data model. In addition to having a dynamic aggregation algorithm at the core of the framework system, this thesis provided a flexible user-experience where the users were able to contribute throughout the phases of the data loading and transformation, analytic execution against the loaded data, and the possibility of expressing the outcome in different visualization representations.

A questionnaire was conducted to further comprehend the significance of the framework system as part of the healthcare setting, and to get feedback from the users after using it for some months. The results of the questionnaire were analysed for statistical correlations, and the outcome was an evaluation of users' experience that targeting different aspects of their engagement with the framework. The outcome was summarized and further work was suggested in order to improve it in future versions.

Co-Authorship

Acknowledgements

Acknowledgements go to my family for support throughout my degree, and in a very particular way to my wife Chantal. Thank you.

Statement of Originality

I hereby certify that all work described within this thesis is the original work of the author. Any published (or unpublished) ideas and/or techniques from the work of others are fully acknowledged in accordance with the standard referencing practices.

(Fadi Louis Nammour)

(12, 2018)

Table of Contents

List of Figures

List of Tables

List of Abbreviations

ANOVA	Analysis of variance
API	Application programming interface
CPOE	Computerized physician order entry
CRUD	Create, read, update, and delete
CSS	Cascading Style Sheets
CWA	Cognitive Work Analysis
DAA	Dynamic Analytical Algorithm
DCD	Decision-Centred Design
DISC	Intervertebral disc
ECG	Electrocardiography
EEG	Electroencephalography
EHR	Electronic health record
EKG	Electrocardiography
EPR	Electronic Patient Records
ERP	Enterprise Resource Planning
ETL	Extract, transform, load
HIS	Hospital information system
HTML	Hypertext Mark-up Language
HTTP	Hypertext Transfer Protocol
JSON	JavaScript Object Notation
KPI	Key performance indicator
PACS	Picture archiving and communication system
PMM	Performance measurement and management
PMMS	Performance Measurement and Management System
POJO	Plain Old Java Object
RDBMS	Relational database management system
SOAP	Simple Object Access Protocol
SPSS	Statistical Package for the Social Sciences
SQL	Structured Query Language
SSM	Soft Systems Methodology
UML	Unified Modelling Language
UMLS	Unified Medical Language System
URL	Universal Resource Locator
VPN	Virtual private network
WBC	White Blood Cells
WCD	Work-Centred Design
XML	Extensible Mark-up Language

Chapter 1

Introduction

1.1 Background of the problem

The evolution of healthcare data is as old as that of human biology. Ever since there was a way to record human health changes, the need to track these changes and record their variations became the incentive for a documented health record of a patient's well-being and illnesses. From that point onwards, healthcare data flow increased in volume and widened in spectrum. The burst of technology and the wide adoption of digital media lead to the wider collaboration between healthcare professionals. Once stored and accessed by one doctor or clinical centre, patient data evolved and changed in identity and content, and through legislation and government-imposed legal, ethical and confidentiality regulations, patient health information because an asset with a single owner and a multitude of access management protocols that controlled who, when and how it is analysed. Patients were no longer restricted to a single point of care provider, but rather adopted the right to change and decide on their care givers. The iconic legacy of the healer-doctor evolved in time to the realistic human care-giver who can make mistakes, and for those mistakes and errors a legal mechanism was adopted to judge performance and responsibility towards human life.

Medical data statistics became more and more available and were used to serve as factual evidence towards criticizing or advocating healthcare plans. With the intertwined nature of sciences and human evolution, healthcare informatics took the stage to methodologically improve the consumption of patient-related data and formulate access to data that is needed by clinicians to deliver optimal patient care.

Interest in healthcare informatics increased with the emergence of the big-data paradigm. Many healthcare organizations had accumulated large volumes of data by that time, and the limitations of the classical client-server data model had been already reached. In addition, the growth of cloud services offered an opportunity for information access and sharing outside the physical premises, creating a new challenge for the client-server architecture of legacy enterprise systems that host patient data. Not only was it difficult to adapt these systems to the limited bandwidth and hardware of connected devices, it was obvious that a new challenge will govern healthcare data-sharing. Unlike other businesses where data producers exposed data based on clearly defined business rules, the healthcare industry lacked the standards that govern the way data is shared, used, and re-used.

In a simple analogy with a sensitive sector such as the banking sector, handling a patient's blood sample extraction for examination differs substantially from a money deposit at a local branch. The two data producers – the blood specimen and the paper money – differ in the way the data operator handles them. A laboratory phlebotomist needs to create a security layer of data segregation when handling the extraction tube. On the other hand, money going into the storage compartment – money drawer in the case of the bank cashier – can be easily mixed with other paper money from other transactions. Furthermore, the electronic transaction that is created to note the deposit of money is immediately validated against the existing business rules. Currency conversion is performed, and the amount is credited to the client's account. All other validations are made, and the transaction is stale in the history of the bank account. On the other hand, in the laboratory system, the transaction itself is another data producer, in that it creates a multiplicity of transactions that need to be performed on various machines, awaiting results to be delivered. Privacy issues arise when requesting, delivering and consuming patient data [4]. In the case of laboratory results, the patient might request that this particular extraction be discretely circulated to his doctor only, shielding it from further investigations that might be done on the patient's medical records. The laboratory centre might have internal safety protocols that disallow release of results of a sensitivity that might be performed against the extracted specimen, while other exam results – say Glucose – might be of a lower sensitivity and automatically cleared for diagnostic investigation without the patient's specific permission. Alarming results are also context related. A machine handling a chemistry result may report an out-of-range result when a reagent is used, while another reagent type might produce a different normal value range for the same blood specimen.

The above example is only one of many that can be used to prove the particularity of healthcare data in terms of privacy, security and sharing permissions. As data producers increase in type, multitude and variance, it becomes a challenging task to keep track of all the sensitivities involved in the healthcare information management process.

Understanding the unique challenges in healthcare informatics is key to improving the performance of care givers and reducing errors that might have fatal effects on humans.

1.2 Motivation of the Thesis

Improving healthcare plans is not an individual task, neither does it depend on individual information. Aggregation and pattern identification is a key factor in decision making in the healthcare environment. Formulating new knowledge based on volumes of patient encounters and diagnosis is a common practice, and in an expanding range of data sources, the challenges are multiplied. Focusing on one data zone at a time to extract temporal data is a common practice solution to handle the enormous job of data analysis. In line of the example provided before, such an approach would be analogous to identifying high glucose patients based on diagnosis. Such studies are common practice in almost every healthcare setting. The problem with this approach is that it excludes – for scoping and performance purposes – other vital relations in the patient data that might otherwise filter out matches that are usually referred to as "false positives". Integrating other data sources in such a scoped study would improve the accuracy of the results but would come at an unaffordable cost in performance and human workload. Information managers in healthcare environments spend serious time defining models and building integration layers that route data from and to information systems.

There are various methodologies for deriving information from data based in a multitude of sources, and the branch of informatics that most relates to this is that of data mining, combined with the technologies developed over the years for data warehousing incorporates knowledge transformed from data silos into optimized databases that focus on the efficiency of information delivery rather than the relational representation of data schemas. Such methodologies were adopted in healthcare settings, creating what is now known as the science of medical informatics. The dependency on medical informatics became apparent as governments, insurance companies and major healthcare players found themselves subject to financial pressure from their adoption of healthcare strategies and coverage plans that were not optimized for the client base. As such, medical informatics uncovered the link between patients' health patterns and the price tag of

hospitalization. Healthcare centres and third-party payers alike were soon investigating new performance indicators such as re-admission rate, medication efficiency, fatality rates, infection control and many others. These performance indicators are directly related to the quality of service influencing the increase in income within a highly competitive domain. On the other side of the table, clinicians were competing to improve patient wellbeing by adopting optimized treatment plans, data-driven best practices for medication efficiency. Clinical performance indicators became central to the public health sciences, and factors such as length-of-stay for patients with chronic diseases and post-operative infections became focus of many studies and published research.
The aim of this thesis is to present a framework of software components that serves as a platform that clinicians use to load and process data, and present findings that can lead to better decisions by stakeholders, financial institutions and insurance companies. The framework is profoundly based on the adoption of performance measurement and management techniques used in other disciplines such as business and engineering. The framework is introduced through a wide literature review that sets the stage for the identification of the need for a new system to serve non-technical clinicians in creating analytical enquiries. Because of the multitude of disciplines that are encountered in such a system design, a focused approach is taken to integrate concepts in a unified system design ([30] and [91]).

Healthcare settings offer a unique challenge to informatics professionals that need to continuously engage a variety of heterogeneous systems that co-exist to deliver patient care services. During personal experience in such a setting for more than 10 years, various scenarios were encountered where data interchange and integration was a major challenge. The requirements and needs of physicians, clinical analysts, stakeholders and external partners do not stop and cannot be under-served.

1.3 Aim and Objectives of the Thesis

The aim of this thesis is to present a software system in the form of a modular framework architecture that allows clinicians and non-technical healthcare specialists to create analytical algorithms to process clinical data produced at different healthcare points of patient encounters.

To achieve this aim, the study proposes the following objectives:

- Examine the current approaches in the healthcare informatics domain, identifying the most similar approaches and critically studying their existing solutions.
- Identify the need for a new approach, based on the limitations of the existing approaches
- Analyse the current approaches and the proposed approach
- Design the new system and then and develop it, following a methodological approach and software development best-practices.
- Evaluate the system and identify its impacts on the environment it is running in.

1.4 Research question and problem

In the process of executing the objectives and working towards achieving them, this thesis will attempt to address a research question, mainly whether the existence of a decision-support framework that allows non-technical clinicians and medical professionals to effectively reach the desires outcomes of their analysis of patient data and health information would improve the performance of healthcare organizations and medical processes that in term lead to an improvement in the quality of healthcare services provided to patients. The idea of allowing non-technical practitioners in the domain of healthcare to perform their study of the continuous flow of patient data in all forms and from different sources, is to be shown in the context of the current work as a solution to the research problem, namely the expensive dependency of medical analytics on technical personnel to drive and deliver analytical results to the people who can use it in their daily processes ([5], [34], [51], [77] and [103]). This problem, if solved, would therefore prove the following hypotheses that this thesis establishes: the existence of a framework that enables non-technical professionals in the medical and clinical domains to achieve a data-centric study of the existing patient health and clinical information in a quantitative approach that delivers results useful in improvement of patient health and safety and can be adapted to realize better results in future analytical studies, such a framework would eliminate the need for technically empowered users to drive medical analytical problems and solutions.

This thesis delivers a novel healthcare data analytical framework that enables clinical and non-technical decision-support officers to achieve a better understanding of the data that comes from different and heterogenous data sources within and outside the healthcare organization. The work is focused around the field of clinical intelligence, which is well-defined in several literature works and has grown out of classical data analytics and became increasingly influential in the healthcare settings due to the maturity of clinical information systems and the acceleration of data flows inside and outside patient points of care. Clinical informatics is defined in [76] as an application of software engineering techniques and tools to clinical data for the purpose of further understanding the different dimensions that this data holds. [76] states that clinical intelligence is "an engineering discipline oriented to the development of methods and tools, to obtain new perspectives of clinical data”. The definition can be further expanded in [68] with the focus on the mechanics of data extraction and the highlighting of decision-support as pas of the outcomes of clinical intelligence. [68] defines clinical intelligence as “a set of computerized methods and processes and the discipline of extracting and transforming data into knowledge that can improve the effectiveness of clinical and health decisions.”

In the process of delivering this framework, the thesis designs and implements a novel algorithm that executes based on a dynamic data model designed for this purpose the analytical steps, aggregations and queries that the users define inside the framework database. The thesis takes the complexity of working with dynamic and changing data models away from the clinicians and users and reduces the need to consult technical specialists when the underlying data sources change, deliver new data models or get altered or retired due to changes in the processes that use them. The need for the new framework system is established after following the soft systems methodology and after conducting a survey to understand the different healthcare challenges and limitations. The survey outcome, once analysed, gives a better visibility of the gaps that exist in the current healthcare environment and the needs of the users when it comes to delivering stakeholder reports and dashboards, where the top management expectations are always changing based on the delivered results. Enabling the decision-support officers and clinicians to communicate their results and findings based on analysing and computing against the healthcare settings data without extensive support from technical personnel is a major challenge in the growing healthcare setting, with more medical devices and systems getting connected and delivering increasing volumes of data to the patient-oriented data stores.

The following diagram shows an overview of how the deliverable framework system is divided into multiple modules and components, and how they relate to each other.

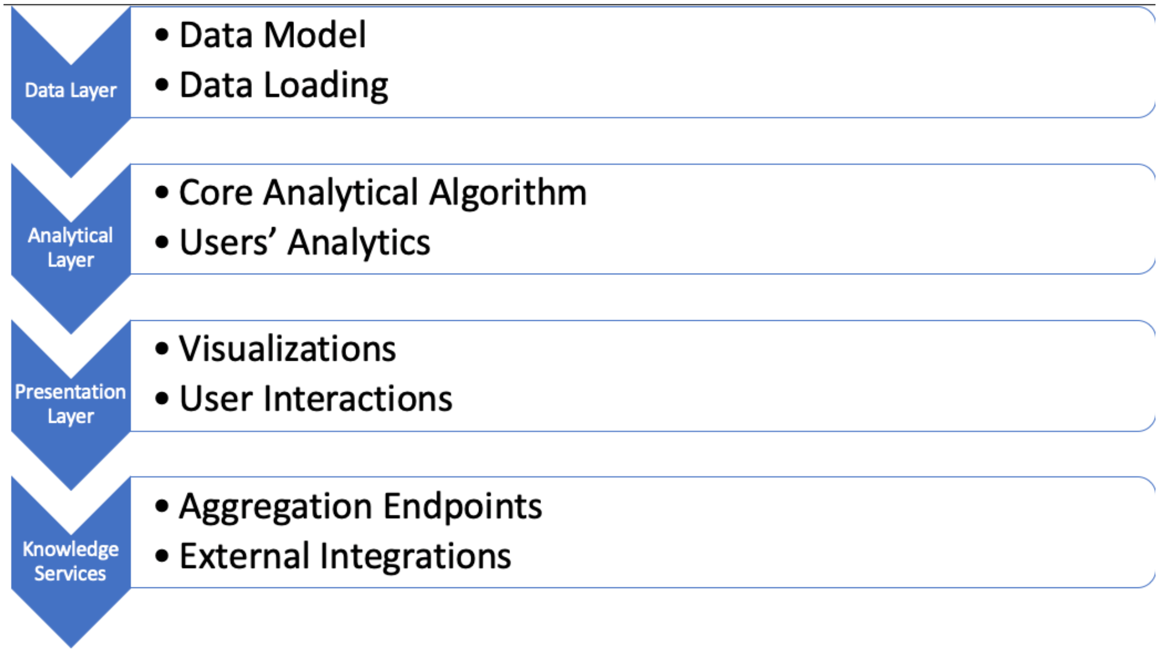

Figure 1.1: Framework System Modules and Components

1.5 Structure of the Thesis

The work in this thesis is achieved in several stages, which are used as milestones in the progress from initial conception of the problem domain to the final evaluation of the outcome. The chapters are designed to progress from a general and global understanding of the problem and current situation in Chapter 1 to a formulation and statement in Chapter 2 of the methodologies that are used in later chapters, followed by a thorough and detailed description of the existing literature in Chapter 3, then deducing the need for a new system in Chapter 4 that overcomes the limitations identified before, followed by an analytical decomposition and examination of the requirements of this new systems and the problems facing its creation in Chapter 5, which establishes the required knowledge and understanding to design and develop this framework system in Chapter 6, which is then finally put to the test as part of an application scenario in a healthcare setting, focusing on the evaluation of this system based on feedback from users in Chapter 7. Chapter 8 concludes this thesis with a summarization of the previous chapters, initiating a self-critical discussion for future enhancements and improvements to the established novelty. An overview of the structure of this thesis is shown in the figure below:

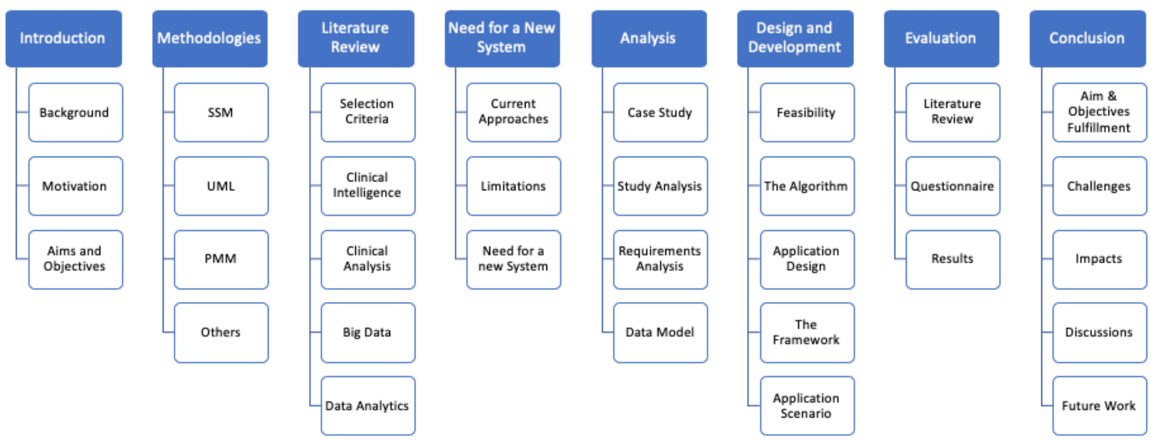

Figure 1.2: Structure of the Thesis

The approach taken to achieve the objectives and reaching the aim of this thesis is also depicted in the diagram below, with a logical flow depiction of the activities performed to reaching these objectives.

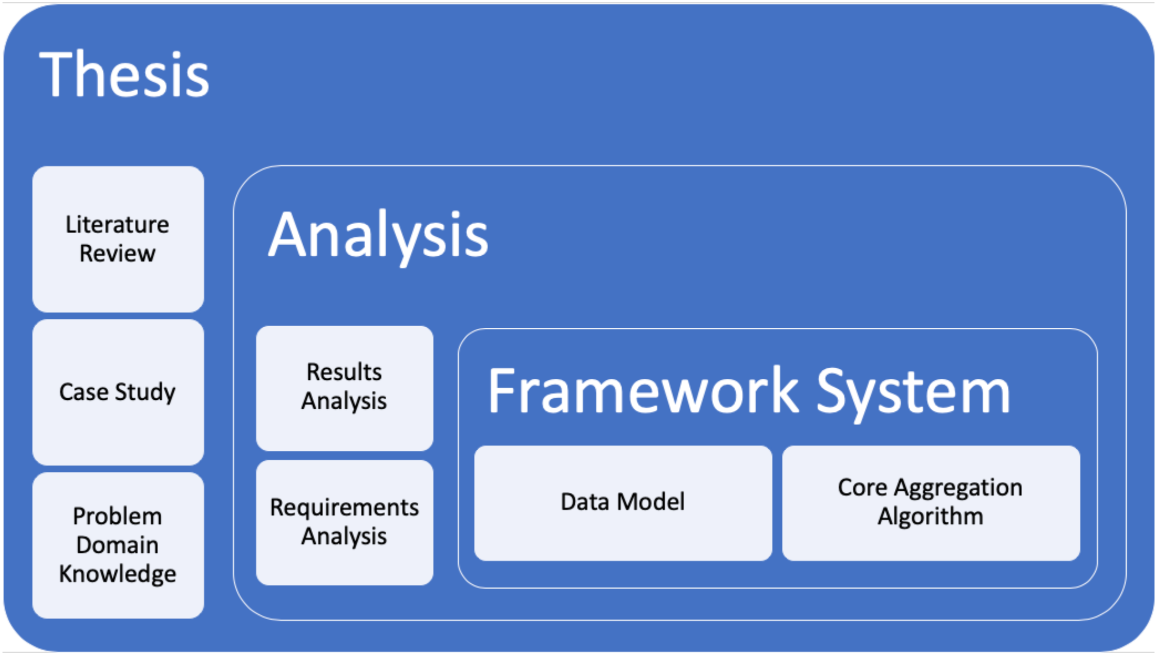

Figure 1.3: Thesis structure and Activities Flow

1.6 Contribution of this Thesis

The aim and objectives section defined what this thesis aims to achieve as part through a set of objectives that are worked against in the following chapters. The work done in this thesis brings a novelty into the healthcare setting through a new approach to healthcare data analytics. A new software system is analysed, designed, developed and tested in a healthcare setting where different data sources form a massive body of medical and organizational knowledge. The framework system achieves this novelty through its ability to interact with the clinicians and the non-technical users in the healthcare domain and guide them through an interactive process to get their own data into the system's data warehouse and then apply their own analytical logic to draw conclusions from the analysis of their own data and from that of other organization processes. The diagram below shows a process flow that users use to create the analytics that contribute to the stakeholders' decision-support. The process is aligned with the framework system's internal modules and component usage, making it easier to follow every stage's internal activities.

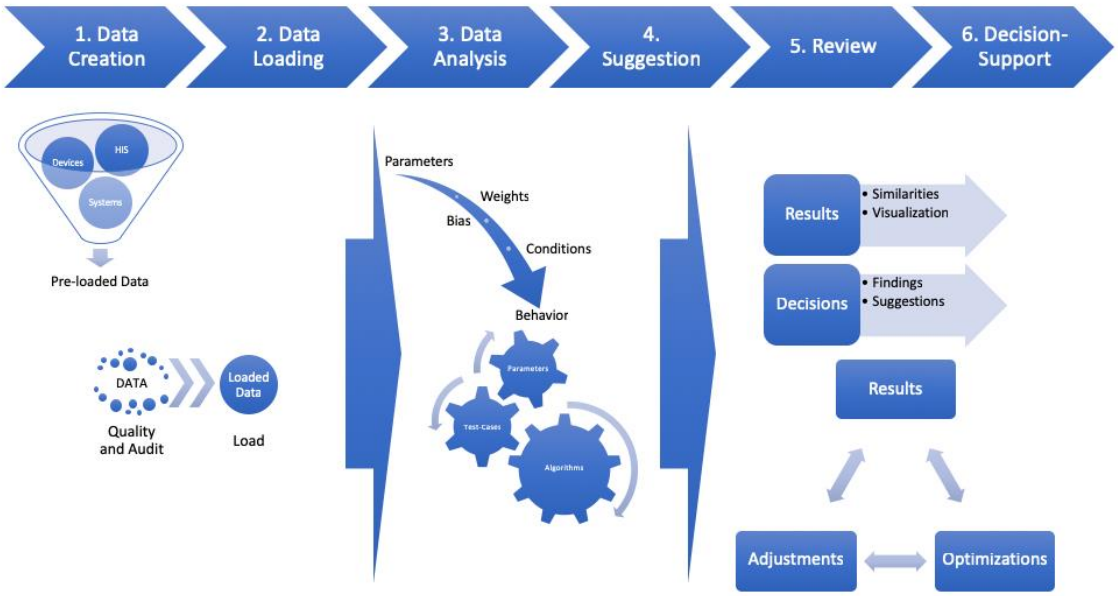

Figure 1.4: Framework system's components process

Another diagram shows a top-view of the framework system's architecture and modular design and is later explained in section 6.6 where the structure of the components is technically detailed.

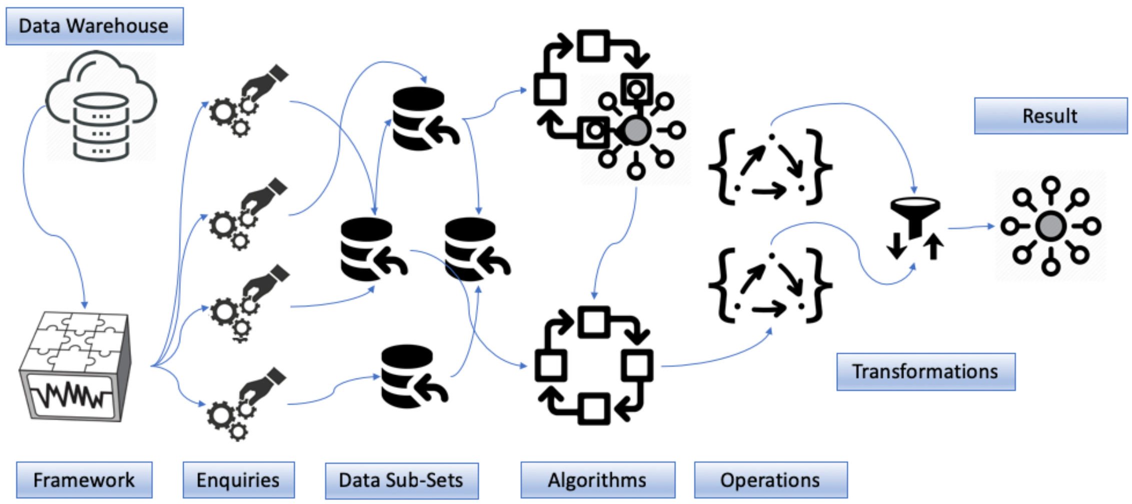

Figure 1.5: Framework system's internal components and modules

The framework system achieves this aim through a new algorithm that collects the dynamic user analytics and input parameters and executes the analytical steps that they defined based on a set of aggregation and grouping functionalities based on the underlying dynamic data model.

1.7 Summary, discussion and conclusion

This chapter presented an introduction to the healthcare informatics domain, where the medical data statistics' availability and usage has served a driving force towards better and improved healthcare plans. Healthcare informatics exists in the context of sciences and human evolution and is therefore in a stage that methodologically advances the consumption of patient-related data and facilitates access to data that forms the basis for building and improving patient health and the quality of service in patient points of care. The accumulated and growing size and volume of healthcare data found in many healthcare organizations could lead to limitations in currently adopted technologies and would therefore exhaust the implementations.

The concentrating on one information zone is a typical practice answer for handling the huge activity of information investigation. There are different systems for getting data from information situated in a large number of sources, and the branch of informatics that most identifies with this is of information mining, joined with the advancements created throughout the years for information warehousing consolidates learning changed from information data sources into upgraded databases that emphasis on the proficiency of data conveyance as opposed to the social portrayal of information patterns. Allowing non-specialized experts in the area of clinical services to play out their investigation of the persistent stream of patient information in all structures and from various sources, is to be studied with regards to the flow function as an answer for the examination issue, specifically the costly reliance of clinical investigation on specialized work force to drive and convey scientific outcomes to the general population who can utilize it in their daily activities.

Chapter 2

Methodologies

2.1 Introduction

This chapter present discusses and explains different methodologies used in this thesis in a manner that best serves the context where the methodologies are used. The use of methodologies to guide the process of reaching an objective is known to be very effective, especially in the study of research related to information systems. When the literature is reviewed for current state-of-the-art approaches, a large range of work is found which would lead to many confusions if reviewed in a linear or non-structured approach. The same is true when performing an analysis of a current problem and determining the limitations in existing solutions. If no structure is followed, it would be difficult to reach the correct arguments and to stay focused on the specific aspects of the problem at hand. The use of methodologies in structured research is key to achieving correct and useful results.

In a search to develop a methodological framework for evaluating clinical processes, [39] performed a study to determine the best approach to look into the evaluation of aspects involved in clinical processes and determined that such activities are distributed over the many patient engagement processes. The work concluded that a process-based approach to methodologically determining a research question is best suited for similar methodological tasks and therefore implied that an activity comparison within this process-based approach would be the most appropriate methodology to apply.

The methodological approach is also studied in [33], where cognitive engineering is put to use as the basis for a selection mechanism that helps find the best ways to present information to humans in a form that emphasises links to ideas and the knowledge that is already acquired. Knowledge capture is found to be best supported through collaborative work analysis and following an application of task analysis methodologies in a structured manner that helps analysts with less experienced users. [33] argues that knowledge is best captured in a methodological approach with a critical decision method in addition to cognitive demand analysis through a structured approach. The work names three cognitive support requirements which are most suited for such methodologies, namely the Cognitive Work Analysis (CWA), the Work-Centred Design (WCD) and the Decision-Centred Design (DCD).

The usage of a family of tools provided by the cognitive analysis approaches leads to developing tailored design methods to improve human understanding of other works and employs meaningful tasks that help users and managers to get the most out of the existing pools of knowledge.

2.2 Literature review of the used methodologies

To review the literature that is relevant to the different research topics described in this thesis, an approach was adopted where many research databases were used to feed the review. The literature review was conducted in a structured approach, which is composed of several levels of information searches and comprehension, followed by a structured listing method that includes categorization of subject material and the development of an understanding that suits the main subject of research that is included in this thesis.

The literature review methodology developed over the time that took for the completion of this thesis and is therefore developed as part of an initial research process stemming from personal knowledge of literature portals and the acquired knowledge of literature categorization, summarization, mapping, and formulating. The first step to generate the literature review, as part of this structured methodology, is to identify the main categories of science that are closely related to the topics discussed in this thesis. To do so, a categorization determination process was undertaken by searching online portals for matching information with a focus on subjects of interest to the understanding of this thesis. The following categories or disciplines were decided and identified, and are Clinical Intelligence, Privacy, Medical Analytics, Visual Analysis, Clinical Analytics, Raw Analytics, Big Data, System Design, Analytic Engine, Clinical Pathway [43], Business Analytics [67], Extraction-Transformation-Loading, Ontology, Usability, Data Analytics, Governance, Evaluation and Methodologies. The topics are related to different levels of importance to the central work in this thesis and are therefore treated differently when used to label or categorize the reviewed literature articles.

After deciding on the categories that are of importance, this thesis moved to identify the most important sources of literature through visiting the most popular online portals, starting from the electronic library search engine provided by Kingston University, the iCAT search engine. This engine is commonly used in most university online portals, which help student researchers to find the information related to their search queries, either in the university library as printed format or digital content, or online in other literature portals, through a unified authentication model which helps redirect searches to a full-text version of the literature work. The iCAT uses a search interface similar to most online literature portals, and the search that was used in this thesis was adapted to this engine. A few guidelines were established to limit the results that would be presented, and also to increase the reliability of the reviewed content. Because this is a study that is based on technology, it was important to limit the age of the search results to those that were published in the last 5 or 7 years at most. This ensured that content of relevance to the researched topics is also up-to-date with modern technological advances. The search was also limited to a peer-reviewed journal published articles. This is important because any content that would be used to identify a novelty in the domains under research would need to carry an acceptable level of credibility that is only ensured when such a publication is reviewed by established review boards.

This thesis also focused on journal publications because conference proceedings and online publications are statements of work that document part of a research process and not the eventually established outcome. Following this search limitation procedure, the iCAT was used to search the literature based on keyword formulation that is relevant to the category under search. A logical query was built to satisfy the completeness requirements of a literature search. The iCAT tool, as other online literature search engines, lists matching literature using a proximity and relevance search method that puts the articles with the closest match at the top of the search results list. This means that when a category or keyword or query is used, the first part of the result list is the one which is of most interest and is, therefore, to be read and analysed. As the proximity to the main search query is exceeded, the results in the lower part of the searched listing are only relevant due to partial satisfaction of the query requirements and therefore are discarded.

For the focus group of categories that was selected before, this also means that a very large list of findings and literature content can be avoided as well, leading to an effective survey of the literature, instead of mixing non-relevant findings with important ones. Following the initial search, the articles were visited in the intent of getting a full-text electronic content which would be saved to local storage for future use. This ensures that lack of access to such a portal in the future, or any other technical hindering factor, would not result in a lack of content available for this thesis. Once downloaded, the article is stored in a single folder with other previously found articles. This is important because, in this way, different queries that lead to the same article would not result in a doubling of content, and in doing so the name of the file that is downloaded is checked for duplicates from previous searches. Another duplicate check is done for the article subject.

The methodology followed in this section is a special application of a wider spectrum of systematic review methodologies. Following the work in ([3], [26], [61] and [36]), a systematic review methodology can be defined as the quantitative and procedural approach where multiple literature works are searched, classified, weighted, and indexed in such a way that establishes a cluster of knowledge-centric published material where the hypotheses and research problems under investigation are validated based on their proximity to the identified truths. The searching and classification techniques used in systematic review methodologies vary from one methodology to another, but they follow a best-practice line of thinking and benefit from the availability of literature in online portals and subscription-based internet services. [3] identifies several biases that are applied during the review searches such as language, publication date and case-sensitive biases. Narrowing down a set of literary works based on well-defined criteria help improve the accuracy of the hypothesis validation and reduces the amount of work that is needed to perform this review. In [26] a systematic review approach is adopted where the occurrences of key-words and phrases are indicators of relevance to the sought subject matter. [26] filter out unrelated articles based on the criteria of abstract relevance and more importantly the availability of full data-sets for their research on the clinical application of polymer gel. This approach reduces irrelevant articles in a more efficient way than that of [3] because it reaches into the content and not just the word-related filtration.

In [61] a systematic review approach is applied to clinical text-based documents to classify them into one of several classes such as pathology reports, radiology reports or medical treatment plans. The classification is achieved through a variation of multiple selections text-based criteria including keyword occurrences, lower- and upper-case words emphasis, tokens and whitespace distributions. The result is a classification method that can be later used with other text-based processing tools to extract knowledge from clinical documents. The systematic review in this work is an application to clinical documents, but it is comparable in its success to the one used in [3].

A modified systematic review approach is applied in [36] to select clinical practice guidelines (CPR's) through a two-stage screening process. The first stage looks for the article's abstract relevance to the investigated CPR before the full-text is further processed in the second screening stage. In addition to abstract relevance, the article's publication date is constrained to only those after 2011. Following the second screening stage, the articles' related links are navigated from the original article and the same systematic review methodology is applied again. The result is a filtered-down list of

CPR's that can be used as a basis for the clinical investigations. [36] offers a procedural systematic review methodology that can be easily adapted to cases where articles are linked to each other, such as paper references.

Following a query search or multiple searches with the same query but in different variations, the downloaded articles are processed in a structured procedure. For every article, a local small database is used to store the name of the downloaded filename, the title of the article, the list of authors – separated by a special delimiter, the category under search, the journal that published the article, the year of publication, a special unique index of the article and the full Oxford citation for the article, and a summary of the article, consisting of key statements in the publication that are of highest importance and would transmit the correct idea. The table structure is shown below, with an example article:

ID	Filename	Title	Authors	Journal	Year	Summary	Topic	Paper Index
1	p734-weiss.pdf	Predictive Rule Discovery from Electronic Health Records	Sholom M. Weiss | Nitin	IHI'10, November 11–12,	2010	learning method is rule	Clinical Intelligen	1

Figure 2.1: The table structure of the articles table

The article summary is generated in an iterative method that is based on an initial contextual reading of the article, where the abstract section is read to determine the ideas that are most central to the article content. Another detailed reading of the article is done, with a focus on the key ideas that are found in the abstract. The paragraphs around these main ideas are especially examined for refining of the article's key statements. Those statements are then reordered in such a way that they convey the article's content in as clear of a sequence as possible. The statements are kept within a count between 5 and 10 statements and saved to the summary column in the articles table.

The methodology used here is custom designed for the purposes of a single literature review, but its usage can be extended to serve multiple reviews of different topics in different works. The methodology can be further improved to make it more systematic through the use of a database of authors, categories, and many other identifying tables. It can also be linked – using application programming interfaces – with other online databases for easier searching and proofing.

2.3 Soft Systems Methodology

The Soft Systems Methodology (SSM) was resulting from work directed at Lancaster University to apply Systems Engineering ways to deal with complex "administration/business issues" [99]. At the end of the day, they endeavoured to apply a Hard Systems way to deal with settle business issues. What they found was the approach regularly staggered at the initial step of issue definition. This happens basically since the diverse partners have different perspectives on what constitutes the framework, the reason for the framework and accordingly the issue. Two key players in the improvement of the SSM are Peter Checkland and Brian Wilson who could put together a down to earth and realistic way to deal with the recognizable proof and arrangement of "delicate" not well-characterized issues. This technique was something beyond a procedure; Checkland and Wilson likewise built up a set of instruments to enable clients to do the means. These include Rich Picture, Conceptual Model, CATWOE, and the Formal Systems Model. Since its root back in the 1970s and 80's it has changed as different specialists have included their own input. Their 7-step process way to deal with SSM allows a few keys and critical parts of SSM to be made clear. It was conceived of the acknowledgement that this present reality is mind confusing and untidy fundamentally since individuals possess it.

Every person has an alternate impression of the same circumstance, and this observation is founded on ideas and convictions held in the mind – a mental model that is used to make judgments about reality. A basic case of this originates from [60] one of the originators of SSM. The author requests that the reader envision two individuals viewing a TV program. Both have observed the very same pictures and heard the very same sounds, yet they achieve totally extraordinary decisions about whether it was a decent or poor program. They express their decisions and afterwards have contention with reference to who is correct. As a matter of fact, both are true in their own right as every one of them had utilized their own particular mental model that involves ideas of what is great or not. On the off chance that these ideas that shape the mental models could be expressed as opposed to understanding, they could be used to analyse against what was watched enabling every spectator to shield their judgment. They may well differ with the particular benefits of their models; however, the contention would now be able to be conveyed out on a saner and more defendable premise. It is the two words “rational” and “dependable” that are vital and frame the premise of the SSM. In basic terms, SSM takes the chaotic contentions of this present reality caused by individuals having unique recognition and makes defendable and level-headed models for correlation with what is going on in this present reality to help made judgments or proposals with regards to the reaction to the issue or issue.

The Conceptual Models are not models of this present reality that are encountered, yet intelligent models of what it could resemble. SSM isn't generally critical thinking in the feeling of investigating this present reality to discover the main drivers of issues. Despite the fact that to the working of the models is the utilization of solid rationale that is derived from an announcement of reason caught in a Root Definition of a pertinent framework. It is vital to take note of that these Theoretical Models will be models of what sensibly should be done to accomplish the reason communicated in the Root Definition. Theoretical Models are a model of what "great" resembles that can be contrasted with reality keeping in mind the end goal to recognize where change could be made. But the dialect of SSM isn't one of consistently utilize. The SSM model starts with the theoretical frameworks thinking, then moves to figuring the Root Definitions, and finally building up the Conceptual Models of what should be finished.

In figuring the SSM, Checkland et al were inspired by frameworks that include people performing assignments and exercises rather than the great gear or machine centred Systems Engineering world that they came. To strengthen that, they talk about the Conceptual Models as Human Activity System, and that is where present reality is captured in the first stage of the SSM, endeavouring to set up what to be considered. The other stage is about what is conveniently going to happen in light of the learning pick-up by looking at the legitimate Conceptual Models against the truth experienced.

The stages are listed below [60]:

1. ***Enter the situation considered problematic***: This stage focuses on current real-world actions and the social event of data and perspectives about circumstances that are considered to be dangerous and, in this way, there is some extension for development. Ordinarily, once it has been concurred that some change or, then again survey is required, this progression additionally includes some essential

research into the circumstance to accumulate data on the key partners and current execution and issues.

2. ***Express the problem situation***: Perceiving that this present reality is muddled, the second step is concerned with catching the different perspectives of the circumstance. To achieve this Checkland et al built up the idea of a Rich Picture to catch the different discernments. They comprehended that complex circumstances couldn't be sufficiently caught by words alone, graphs and pictures are far more viable and can pack a higher thickness of data per square centimetre. The thought behind the development of a Rich Picture of a specific circumstance is that it allows contrasts of translation to be distinguished, permits consent to be made on the translation to be taken, and is a wellspring of motivation with respect to what important frameworks could be demonstrated through the absorption of connections, issues and so on. It distinguishes subjects to take into the framework's world. Since each circumstance is unique and it is important to catch this potential assortment, there are no formal Rich Picture displaying images. Be that as it may, throughout the times of utilization, a number have progressed toward becoming acknowledged as standard.

3. ***Formulate Root Definitions of relevant systems of purposeful behaviour***: This is a basic venture in the SSM. The Root Definition is a mission statement that catches the embodiment of the specific circumstance of the applicable framework. At the core of the Root Definition is the change that is performed by the applicable framework. This is caught by the principle verb in the Root Definition. The Root Definition is critical since it is this that is utilized to intelligently derive

what the organization will need to do all together meet the definition. This is caught as a Conceptual Model. To help guarantee that a draft Root Definition is worthy Checkland and Smyth (1976) built up the mental aide CATWOE where:

a. The Customer: The individual(s) who get the yield from the change (in late circumstances it has been perceived that the out of the change might be "negative" for a few clients and "positive" others. This has prompted a refinement of CATWOE to BATWOVE where the C is broken into Beneficiaries and Victims!
b. The Actors: Those people who might DO the exercises of the change if the framework were made genuine.
c. The Transformation: The intentional action communicated as a change of info to yield.
d. Weltanschauung: It's a German word that actually signifies "world view". It is the conviction that comprehends the root definition.
e. Owner: the more extensive framework chief who is responsible for the execution of the framework.
f. Environmental Constraints: the key compels outside the framework limit that are critical to the framework.

4. ***Build conceptual models of human activity systems***: Checkland recommended going for 7±2 exercises that are at a similar scale. These are frameworks that accomplish their motivation through human action rather than programming escalated frameworks or equipment serious frameworks. The operational exercises are those exercises that are legitimately important to play out the change

expressed in the Root Definition Each movement should be checked to decide if it is being done well and control move made on the off chance that it isn't. The main observing and controlling the operational exercises would address Efficiency and Efficacy. The second level is observing and controlling the checked and controlled operational exercises. Right off the bat, he contended that each action in the operational subsystem should be checked and consequently there ought to be a bolt from the different exercises to the "Screen Operational Activities" action. Every one of these bolts would add superfluous detail and prompt disarray; thus, Wilson acquainted the square bolt with speak to execution data gathered from each operational action. Similarly, the "Make Control Move" action can create a control yield to each operational action. The key thing to recall is that that a sort out is being attempting, with a Conceptual Model, a coherent perspective of the exercises important to accomplish a framework reason.

5. ***Compare models with real-world***: is the place where the present reality is returned to and contrasted with the truth encountered and that caught in the models. The reason for the examination is to start exchange from which changes to enhance the circumstance can be distinguished. The approach utilizes the models to give a method for seeing an alternate perspective of reality by testing presumptions that may exist yet are badly established. It is the contrasts between what occurred as a general rule and the sensible model that brings up the issues that will at last prompt change.
6. ***Define changes that are both desirable and feasible***: In the perfect world, every one of the proposals would be actualized. The organization is nonetheless, a live

association with limited assets (as far as individuals furthermore, cash). Evenmindedly one would have picked the request and timescales for actualizing (or not) the proposals. While this sounds like a simple advance, it is in certainty loaded with troublesomely. Individuals would not generally be inspired to execute change regardless of the possibility that it is established on the rationale of the Theoretical Model. Since SSM was produced for Human Activity Systems it is important to perceive that individuals associated with the potential change could hold clashing perspectives regardless of the possibility that the rationale of the Conceptual Model is certain. On the off chance that change and culture conflict – culture wins. This requirement for social plausibility is something which researchers and specialists now and again find troublesome. They tend to overemphasize the significance of rationale, and neglect to see social viewpoints that decide if or not change will happen. This is one motivation behind why it is vital to contemplate the Weltanschauung of each Root Definition.

7. ***Take action to improve the problem situation:*** When the progressions that are considered desirable and feasible have been distinguished, exertion is exhausted to actualize these. This usage will result in new frameworks that will influence the greater framework prompting more open doors and issues; thus, the procedure begins once more.

2.4 The Unified Modelling Language

The UML is proper for demonstrating frameworks going from big business data frameworks to appropriated Web-based applications and even to hard continuous inserted frameworks [8]. Figuring out how to apply the UML adequately begins with framing a calculated model of the dialect, which requires learning three noteworthy components: the UML's fundamental building hinders, the tenets that manage how these building pieces might be assembled, and some basic instruments that apply all through the dialect. A displaying dialect is a dialect whose vocabulary and standards concentrate on the reasonable and physical portrayal of a framework.

UML is composed of three building blocks.

1. ***Things:*** Things could be structural, behavioural, grouping, and annotations.
2. ***Relationships:*** A Relationship is usually a Dependency, an Association, a Generalization, or a Realization.
3. ***Diagrams:*** Diagrams envision a framework from alternate points of view, so an outline is a projection into a framework. For everything except the paltriest frameworks, a diagram speaks to an omitted perspective of the components that make up a framework. Diagrams could be Class, Object, Use case, Sequence, Collaboration, State-chart, Activity, Component, or Deployment.

Distinctive stakeholders end clients, investigators, engineers, framework integrators, analysers, specialized scholars, and venture managers each convey diverse motivation to a venture and take a gander at that framework in various courses at various times over the task's life.

There are three sorts of connections that are particularly essential: dependencies, which speak to utilizing connections among classes; generalizations, which interface summed up classes to their specializations; what's more, associations, which speak to basic connections among objects. Each of these connections gives an alternate method for consolidating reflections.

To demonstrate a straightforward Software Application [95], a few stages are required as direction. It can be alluded to Agile Methodology to make the means for demonstrating the Software Application as straightforward as could be allowed, simple to learn and for best performance usage to satisfy client's necessities. For a specialist, the study will turn into the alternate route information to show the product application. After gathering the client necessities to fulfil, in light of perception reports and meeting stories, at that point, the information will be displayed as a database.

As for user requirements, it is not possible to know the necessities, desire, long for the client in the event that the client prerequisites are not broken down. Client prerequisite will be done keeping in mind the end goal to fulfil the client. For client prerequisites, it is not required to bother with UML outline, yet normal client necessity strategies like meeting, perception, records, study, Joint Application Development and different systems are adequate. For the learner, the initial three procedures like meeting, perception and record confirmation must be done to think about the information and experience of the client. Meeting and perception systems can be utilized in non-formal circumstances and without documentation while getting the data from the client and so as to fulfil the client. Individuals can fabricate the framework without doing client prerequisite methods, yet it is smarter to utilize them. On the off chance that the framework is fabricated then it would be fulfilled to the client's requirements, and top-notch programming that meets the client needs is created, to fulfil the client than at any rate needed to chat with them, in a non-formal way and in a compelling way.

A product development strategy is measured in the event that it enables the designer to create programming frameworks made of self-ruling components associated by a lucid, basic structure [102]. For the programming engineer, the standards and the rules are similar as critical as the criteria. The distinction is essentially one of causality: the criteria are commonly autonomous - and it is for sure feasible for a technique to fulfil one of them while damaging a portion of the others - through the tenets take after from the criteria and the standards take after from the guidelines. Important to consider are the five basic satisfaction criteria: Decomposability, Composability, Understand-ability, Continuity, and Protection.

2.5 Performance Measurement and Management System

The Performance Measurement and Management System (PMMS) outfits organization with the instruments and the structure by which three fundamental limits are engaged: Communication, Information, and Control [32]. The PMMS perceives lacks in execution and districts that need intercession and change. In an unfaltering space, changes to the PMMS bring about obvious changes in the devotion of benefits. The thought of overseeing hierarchical execution is in effect generally acknowledged and embraced everywhere throughout the world. Execution is alluded to as being tied in with taking the necessary steps, and additionally being about the outcomes accomplished. Execution Measurement must be considered as a component of the generally Execution Management framework and can be seen as the procedure of evaluating the productivity and adequacy of activities.

The most punctual foundations of Performance Management can be followed back to the utilization of money-saving advantage examination in the 1960s; to administration by destinations in the 1970s, and to yield planning in the 1960s. All through the 1960s and the 1970s, a long interim of very nearly twenty years much speaks occurred about bringing Performance Management into the general population segment. Towards the finish of the 1980s, numerous frameworks of Performance Management were conceived, embraced and actualized at many levels of the general population segment. While trying to conquer these reactions, Performance Management systems have been created to empower a more adjusted view amongst inward and outside elements, budgetary and non-money related measures. These Multi-dimensional structures concentrated more on non-money related data and are intended to give an adjust by including measures of outside progress and inside execution, and measures which are intended to give an early sign of future business execution and a record of what has been accomplished before. Progressively, the quality measurement turned into a fundamental piece of most, if not all, Performance Management frameworks. The US's Reinventing Government is organized around numerous standards vigorously identified with Performance Management with a solid concentrate on nature of administrations.

2.6 Evaluation Methodology

Every information system (IS) that is built to improve, advise or develop one or more real-life business processes needs to be subjected to some kind of evaluation that helps measure what it promises to achieve against what the end-users eventually get out of it. In many recent works, the subject of software evaluation has taken considerable attention, particularly due to the expensive and cumbersome efforts that are consumed when the planned solution is not matching the delivered one. In [1], it is found that the evaluation of such systems is useful in determining the end user's acceptance of such systems, showing how efficient and accurate information systems are, how useful they are, and both the features that make such systems powerful and incapable, leading to a quick or slow development in the system's growth. The work reviews a number of papers that researched the evaluation of healthcare information systems, focusing on hospital and health institution research for electronic medical and health records in Iran. The findings show that the information systems affect different aspects of a healthcare environment in different ways, with the highest effect on data quality, then on data management and then on the usability of systems. The work in [1] noted that several methods for information systems' evaluations are being used, with a special focus on the effect of such systems on the medical and administrative procedures. The large adoption of commercial information systems without a standard set of concepts between development companies makes it difficult to arrange for proper evaluations of the effects of such systems in healthcare environments, but the work also notes that in other countries and on a large scale, the adoption of information systems in a healthcare environment has helped resolve many issues, particularly in decision support and computerized physician order entry (CPOE). The authors indicate, however, that any failures or errors in identifying the weaknesses or deficiencies in the studies that they explored could be the result of getting the hospital employees' viewpoints, as compared to a completely objective approach that might have resulted in more accurate evaluations. Another approach was taken in [84] where the authors discovered during their survey of successful implementations on hospital information systems, that many such implementations resulted in failures, which they could compile into a set of success factors that help guide any future implementations. The authors found two evaluation methods, with a different view of how the HIS is implemented. The first method treats the HIS as a sequential process where stages follow a linear model with several defined variables at the level of the organization which lead to a scale of adoption that can be measured during evaluation. This method's approach lacks the ability to adapt to complex processes that exist in any healthcare setting and would fail to correctly evaluate IS components that are built to follow such changing processes. For example, the software module that implements the process of ordering drugs from the pharmacy would not be accurately evaluated using this model, as then the linear approach would lead to a drug fulfilment process that never changes over time, which is clearly not always the case. Any adaptation to the process would lead to confusing this evaluation model, leading to false devaluations of HIS component in question. The second more complex approach that [84] mentions is based on iterative and multidimensional process where the evaluation model would be sensitive to changes in the implementation stages, giving the evaluator the possibility to improve the accuracy of evaluation as many factors such as technology, business expansion, adoption of new sub-processes or even a change in the approach of doing business would trigger better results in the organization. Also, in the scope of the evaluation methodology is the on-site training and user-support efforts of the HIS, which are also marked as critical factors to the implementation success. The factors identified are not only specific to a healthcare setting but also are internationally found to lead to a successful implementation of HIS. The involvement of the end-users is also found to be very critical in the implementation, and therefore users should be engaged in implementation aspects related to acceptance of design and testing of system outputs.

It is therefore important to select a combined approach for IS evaluation methodology in a healthcare setting, due to the complex processes that co-exist in such a heterogeneous environment, and also due to the many factors surrounding an information system that delivers decision-support, namely the human factors that contribute to its success. The people who are using, providing and consuming the system inputs and outputs are clearly part of the evaluation mechanism, and their positive contribution is a major contributor to the successful system that they utilize.

The evaluation methodology is chosen to integrate a technical evaluation of the system components under study, in particular, the modules that it consists and the level integration that it shows when dealing with complex problem scenarios. The evaluation methodology would also focus on the human factor involved in the system evaluation and would, therefore, collect information relevant to the user-base in focus and would range over the many user-groups that would be involved in such a system usage. Both the technical and user-based evaluations would then be combined to yield a global evaluation of the system, which serves to show its elements of success and its weak points that need to be improved. User feedback and output evaluation are both used to feed back into the system building model to improve it in future versions and clear any potential failures that could arise.

2.7 Methodologies Relevance to Objectives Achievement

The different methodologies that this chapter explains serve as guidance for the work in the next chapters. Every methodology is selected based on the existing literature and the best practices that others have adopted over years of research and fine-tuning. The relevance of the methodologies to the objectives' achievement is understood through an examination of each objective and which methodology is used in this thesis to explore and achieve it.

The first objective of the thesis is to examine the current approaches in healthcare informatics and identify the similar ones, studying them critically. The approach is achieved through using the literature review methodology from Section 2.2, where the best practices learned from the literature reviewed shows that it is very efficient to breakdown the massive literature content into categorized sub-fields and index the works in such a way that they can later be grouped into relevant sets for critical comparison. The second objective is to identify the limitations of the current approaches and to achieve that the approaches from the literature review are broken down into features that are then tabulated and explained through the literature review analytical methods from Section 2.2. Once the limitations are identified, the thesis addresses the third objective which is to analyse the current approaches and the proposed solution. The purpose is achieved using the soft systems methodology from Section 2.3 and applies the SSM procedure to extract real-world information and requirements and bring them back to the analytical rigid domain to understand them and solve the problems at hand, and then take the solution back into the real world and apply it to the current situation.

The next objective is to design and develop the software framework system, and for that, the methodologies of UML and rapid software development from Section 2.4 are applied to create a data model and describe the system components and modules in a clear way. The components and structure of the proposed system are developed by applying the best practices of software engineering, resulting in a coherent and modular framework. The last purpose is to evaluate the framework system and to achieve that the thesis applies the framework system in a healthcare setting where users are trained to adopt the new approach in their daily works and use it to deliver comprehensive and descriptive information to the stakeholders and top management. The analytical components of the framework system are used to feed a monthly status report portal with aggregations from the different data sources in the healthcare setting. This application scenario is evaluated using evaluation methods from Section 2.6 and the results are presented in the evaluation chapter.

2.8 Summary, discussion and conclusion

In this chapter, several methodologies were reviewed in different domains of relevance to the objectives that were set in the previous chapter, and the different methodologies were compared to find the ones that most suit the objectives in focus. In the next chapters, those methodologies will be put to use and will be followed step-by-step to reach scientifically-proven results leading to the successful addition to the existing body of knowledge.

Scientific works that aim at improving the existing situation in any discipline need to first study the existing complex and heterogeneous surrounding by first looking into other works that contribute or lead to the subject being researched, and would then identify the gaps that exist, focusing on how the new improvement could yield a better state and improve the existing situation. In doing so, research should use methodologies that are based on solid scientific and proven systematic working techniques to achieve each of the objectives that it sets. Without such a proven methodology, a researcher could lose the focus and delve into an endless maze of trials to find the route with best results. The need for a methodological approach in every scientific research is shown to be of great importance especially when it comes to delivering a concise and readable work that documents years' effort of experimentation and trials.

Chapter 3

Literature review

3.1 Introduction

This chapter presents a review of literature produced in various fields of science and medicine that contributed to the build-up of the need for clinical intelligence in the point of care. Recent journal publications set the foundations for solid research and open the possibilities to investigate problems faced by practitioners and clinical analysts. A structured approach will be taken, considering the vast array of published material and selecting a narrowed-down range of references that help in outlining the roadmap leading to the statements of problems faced by healthcare professionals. The chapter first starts with a review of research published by year by keyword, based on searches in portals and online libraries. This chapter categorizes the publications based on relevance to a subject-matter, and they are grouped by the suggested category. Then the different categories are approached, comparing work done and the year it was published. This gives each subject a time-based perspective, and therefore linking subjects into a time-based evolution schema, which should serve as a solid introduction to the current research. This approach is not conclusive, and there might be publications that were not. However, the approach was developed using esteemed research portals and should reflect a true status of the research community.

A seven-year span was used to filter publications, with a focus on journals that have the most relevance to the main topic of clinical informatics and intelligence. With a scoped approach, search results are analysed to make sure that the search keywords are present in the abstract or at least relevant to the introduction and conclusion sections of the publication. The weight of the keywords in the content of the publications was not emphasized because it is often the case that the authors use variant taxonomy to describe the different aspects of the material being researched. No bias was used to affect the selection criteria in terms of countries, research centres, or research methodologies.

3.2 Selection of literature works

The sources utilized in this literature review include online portals, university libraries and journal search engines. Sometimes the publication would show on different keyword searches, and for those cases, they are labelled with the topic that most describes their content.

Labelling a paper with one topic is a very difficult choice, and for that, a mapping schema had to be developed where the topic was derived based on relevance to key ideas that are important to the domain of work.

A tabular matrix for holding this mapping is displayed below. The acronyms used are CI (Clinical Intelligence), BD (Big Data), DA (Data Analysis), MA (Medical Analytics), CA (Clinical Analytics), MI (Medical Intelligence), and P (Privacy).

	Intelligence	Clinical	Medical	Data	Analytics	Privacy
Intelligence	-	CI	MI	BD	DA	P
Clinical	CI	-	-	CA	CA	P
Medical	MI	-	-	MA	MA	P
Data	DA	CA	MA	BD	DA	P

Analytics	DA	CA	MA	DA	DA	P
Privacy	P	P	P	P	P	P

Table 3.1: Literature Topic Distribution

When a publication is examined, it is determined how relevant it is to the basic concepts listed above – both in the rows and columns. The top 2 relevance ideas are captured, and - based on the matrix – the paper is identified with a general topic. As an example, a paper might include primary information about data-related activities in the clinical domain, so in this case, this paper is labelled with the "Clinical Analytics" topic.

3.3 Discussion of the literature works

The selection methodology helped focus the comparative argumentation for papers within the same scope of work. It is necessary to outline the topics based on relevance to the central topic of clinical intelligence, so each of the remaining topics was linked to how essential it is in relation to the central topic. For example, medical analytics was identified as more relevant than big data, because the topics are focused around the point-of-care and how the caregiver can improve the quality of services provided to improve the health of the patient. It does no undervalue the importance of big data to the medical and clinical fields, however, but it does express the higher value of medical analytics.

3.3.1 Clinical Intelligence

Clinical Intelligence is based on the utilization of artificial intelligence techniques in clinical data analysis. The primary technique that is most frequently used is fact prediction, where information in the form of text, data values, or relationships is input to predictive engines and conclusions are made based on parametric conditions. In this approach, information is consumed from underlying information systems or is provided in the form of documents to be processed, such as EKG graphs or x-ray images. No matter in what format the input data is, it is the responsibility of the prediction algorithm to make sure that it is converted to predicate statements that can be consumed and analysed. In [97] the prediction algorithm targets literature accessible in the healthcare setting, and the output of the predictive rules is a set of conclusions, which can be positive or negative. Correction of the conclusion is fed back into the rule engine for optimization purposes. The authors claim to obtain better prediction results and focus on the ability of the predictive engine to express the predictions using medical terms. The strength of this approach is a structured input-set, where medical content is tabularized and consumed as input to the prediction rule engine. Such standardization of input is key to achieving correct predictions or analysis in the case of medical information, where data is not structured in its raw format.

It is not always possible to tabularize medical information, as is clearly the case in [82], where image processing is on the basis of the prediction engine. Here a user-assisted process is put to work, where radiology images are analysed through a regional comparison technique called “temporal subtraction”. This involves converting the images to a pixel matrix, where the images are subjected to a time-based pixel comparison process. A matrix subtraction algorithm generates “facts” that are then applied to the images and validated by the radiologist. This prediction algorithm is scoped by radiology imaging but is clearly comparable to the approach used in [97]. Aligning the input matrix in [82] with the input table in [97], it is possible to conclude that a tabular-analysis algorithm is common to both approaches. The accuracy of the prediction is subjective in the case of [82], where a radiologist applies a reading to the output. In the approach of [97] the information produced is validated with a clinician's analysis of the medical record.

A third approach for medical prediction is explained in [93]. The authors attempt to apply a special scheme of regression analysis in the form of hierarchical analysis of input from "subjects". The input is a scaled-variable that is based on statements to be evaluated by users in a survey-like input gathering process. The approach is domain-specific, and it deals with the cultural prediction of cross-cultural adaptation issues. This is particularly of no direct importance to the scope of the review, but rather the procedure of handling input – a user survey in this case – in a predictive model as part of an intelligence structure.

Another form of drug-related clinical intelligence is examined in [89], where drug metabolites and toxicity are predicted based on similarity and previously encountered drug sequences. A data warehouse is constructed to host thousands of drug and drug-similar chemical structures, where the input is the chemical composition of the drug and the output is the level of metabolites in any given drug structure. The method by which data is analysed is based on similarity analysis, which means that historical patterns are searched for the most similar and based on which the outcome is predicted. Such a model is a basic pattern-analysis algorithm, but the authors do not clearly indicate how such a model compares with more complex analytical models. From a usability perspective, it is easy for the pharmacist or drug scientist to evaluate outcomes through an interactive system.

A text-based analytical engine is described in [15], where notes in medical records are searched for matching patterns in UMLS terminological. This search is conducted on a language-specific basis, but the logic can be easily generalized. The concept and implementation serves as a presentation of contextual search algorithms, often needed when information is presented in textual format. The UMLS terminological serves as an international database for clinical terminologies, and thus it is important that any medical text describing patient status and health should correspond directly to generally-accepted terminologies. A variant of MetaMap – a system designed to achieve a similar task using English-based terminologies – is adapted to Italian and is applied to the input notes in medical files. The steps of MetaMap are tokenization, parsing, variant generation, candidate retrieval, candidate evaluation, and mapping construction. These steps describe the process of consuming, analysing and identification for contextual formats. The user input is usually a paragraph describing a specific medical condition, and the system attempts to construct the mapping to UMLS terms. This text-to-terminology mapping is critical to transforming free-text to usable data inputs into another system that supports clinical decisions. Although the concept in [15] is an adaptation of an existing algorithm, it embodies an algorithm that proves the possibility of using text-based user input as concrete data input to analytical algorithms. The results are analysed by clinicians to detect missed mappings, and thus a correction is made that eventually leads to a sound result.

A focused approach to provide clinical intelligence through modern information analysis techniques is described in [85]. As patient information is available in different formats, both structured and unstructured, it is important that tools like the one presented in [15] serve as input analysis and preparation stages for a clinical intelligence system. Artificial intelligence tools are extended to analyse captured data and provide guidance in the form of decision-support information. Knowledge engineering is also utilized to construct the input data, and immediate application of artificial intelligence results in turning medical data into smart data, which is the basis for accurate decision-support. Along with statistical information gathered through clinician studies and analysis, a machine-learning model is constructed and applied to facilitate the consumption and correction of such data to produce valuable results.

Another important application of decision-support and diagnosis-prediction algorithms is the continuous monitoring of vital signs. When high-volume input is produced that requires immediate interventions from the side of doctors, it is important to have a velocity-optimized data-mining algorithm that consumes and streams vital sign decision-support directives to the observing clinician. In [27] this approach is emphasized, and incorporation of various data tracking and analysis techniques is optimized to handle the flux of vital signs. The output is a continues time-series which are then fed to monitoring tools. With massive information coming from tools such as ECG and EEG monitors, the task of processing and prediction cannot be efficiently achieved through classical data-mining algorithms. The work in [27] combines the techniques that have been optimized over periods of time in the big-data domains and delivers cloud-based analytics to observing clinicians. Patient data safety and security concerns are well-addressed, and a careful security validation of input channels is maintained through standards of cloud-based security and data exchange. The results of streaming of time-series data are also consumed by remote clinicians, and that also is carefully secured.

It is important to note that in most situations where clinical information is acquired through heterogeneous systems, a combination of adaptations of two or more algorithms is required. Therefore, it is critical to decision-support end-users that an accurate output is achieved, and research is focused on optimizing intelligence algorithms to serve this end.

3.3.2 Clinical Analysis

In the early stages of consuming data from healthcare systems, an important aspect of data screening is emphasized. Forming error-free data inputs and producing accurate output is vital to the quality of successful decision-support algorithms. Clinical analysis research has improved recently due to demands for fast and responsive cloud-based analysis tools.

The sensitivity of patient information that is used in classroom and other study locations is addressed in [35], where information from text sources is extracted and represented in the form of XML documents, and later fed to relational analysis tools for prediction and decision-support purposes. The information that is initially located in documents is later reproduced to structural XML formats that make analysis easier. However, there is no attempt to optimize the analysis techniques once the input data is structured. The algorithm uses techniques such as template identification and code-extraction to formulate a structured XML document that serves as an equivalent representation of the initial document.

A statistical classification approach is presented in [49], where predictive analysis is produced by the application of pattern analysis to a relatively small dataset with a limited attribute range. The focus of the work is to produce a predictive model where weights are attached to attribute values and compared to outcomes from previous similar patterns. By adapting the weights, the nodes can predict the outcome based on historical patterns. This statistical analytical approach does not guarantee accurate results when expanded to larger datasets with dynamic attributes, as is the case with most modern clinical systems, but it does utilize proven statistical methods and serves as a validation mechanism when the data scope is limited. Clinicians can also benefit from common statistical skillsets to support decisions. In a typical healthcare setting, there are some practical applications where the volume of data is not so large. For example, a variation can be applied to the problem of predicting the correct antibiotic for pregnant women with infections. The error function is the extension in the length of stay, and the weights can be the dosage and the treatment duration.

Another similar statistical analysis technique is found in [44], where the main problem of screening emergency patient admissions is examined. ED patients are classified using a scale system, and the length of stay is analysed per category. The outcomes of being treated, referred to another hospital or diverted patients are mapped to input categories, and the study attempts to minimize the length-of-stay for referred patients that can be otherwise treated at another smaller hospital. Admission ratios for ED patients is a classical KPI and are studies repeatedly to reduce false admissions and minimize erroneous diagnosis. The value of this work from a decision-support perspective is that it proves the possibility of generating useful recommendations using classical data processing tools. It is not clear, however, whether the same techniques and tools could efficiently yield accurate results when the data scale is larger and more complicated. It is not uncommon to see similar tools at work in various study situations.

Another analytical approach is taken in [64], but this time a software system is designed to assist in the pattern selection process. The work attempts to predict the most appropriate treatment plan given a historical set of similar patient encounters. The system is fed information from laboratory results, medications, symptoms and diagnoses, and generates a list of best-plans for application in similar situations. The authors' main statement is the accuracy of the pattern filtration functionality compared to other similar systems. In a pattern-matching prediction model like this one, the accurate screening of historical patterns is critical to the successful prediction. Treatment plan selection is yet another challenge to clinical intelligence systems, and the application of analytical techniques is successful only when the number of variables is small. Other factors may be involved in the success of a treatment plan, and once the system is outset with fixated variables then it is not clear how new variables can be later applied to reduce erroneous selections. The work – however – is a step forward from classical analytical techniques and tools, and clearly serves the purpose of arguing for the limitation of prediction techniques when using classical statistical tools.

A moderate application of statistical analysis in clinical studies is outlined in [100]. An association model is presented where prediction of future diagnosis is performed using earlier symptoms and behavioural patterns. The analysis is performed using classical tools and is applied to a relatively small dataset of cases. The study methodology is characterized by the introduction of a control variable – the gender. This is some improvement over other analytical methods where the variables are fixated beforehand, and the results are predicted using a variety of weights. The control variable functions as a moderator which focuses the results in a specific decision path. The variation of the decision based on the late-introduction of a moderator is a key factor for improving decision-support. The work does not establish a generalization of the applied methodology but rather limits the results to the scope of the current data types, namely loneliness and the effects of anxiety and depression to adult life stages.

The limitation of analytical techniques presented beforehand was always the predefined set of variables in the scope of the input data. This means that data is collected in a pre-defined format that does not change over the lifetime of the study. The work in [18] attempts to bypass this classical limitation by developing a meta-data library of classification terms that are "labelled" to incoming data based on a pre-defined mapping of keywords and terms. This method is tested with an inflow of data from various data sources in different schemas. The input data is mapped, and a scoring table is constructed for the various terms defined in the meta-database. Based on the score of each term, a prediction with the most likely association between the input dataset and the highest scoring terms is generated. The accuracy of the prediction is not the primary objective, but rather it is the presentation of a guideline for the observing clinicians that for a given input set, the scope of data is likely limited to a few terms. This type of classification technique can be useful when screening data from different domains into a centralized study repository.

A sampling of data is further analysed in the work presented in [62] and [45]. The emphasis is on exploring relationships between variables in data samples and the corresponding characteristics of the subjects of the study. Both present an adaptation of classical statistical correlation, regardless of the technology used. The predictive models based on statistical correlation is greatly employed in many numerical analysis domains, and thus it is also of value for basic analysis in clinical setups. When the structure of input data is well-defined and data formats do not vary over the period of the study, it can be processed with statistical tools to produce a predictive model. The accuracy and reliability of the results depend on the consistency of the input when applied to current situations. When input data formats and affected variables can be guaranteed to be consistent and when the effect of other non-included variables is minimal, then the output of the statistical analysis of a moderate input-set can be generalized and used for decision-support for larger samples.

3.3.3 Big Data

The growth of the healthcare sector, along with the accumulation over time of patient data, has served as a primary application for the methods found in the recently growing field of Big Data ([20], [23], [24], [50], [54] and [63]). Not only did data grow, it also started including new formats and fact-recording techniques. The application of big data to medical and clinical analysis served as a good improvement in the overall quality of decision-support.

In [74] this variability and diversity of data formats is studies and related to big data models. The work focuses on a multitude of data formats including the classical structured records, multimedia formats such as images, video recordings of operations, and modern social-based records such as posts and communications from social media sources. The work highlights the importance of big data in healthcare settings, with a special focus on public data in patient records. It shows that the organized information in EMRs and EHRs incorporate well-known information record fields and that the data is effortlessly coded into and taken care of via automated databases. The work emphasizes the fact that while the accessible systems and apparatuses are for the most part open source and wrapped around Hadoop and related stages, there are various tools that engineers and clients of huge information examination in human services must consider.

A similar approach is taken in [92] but is extended to incorporate healthcare data in big data tools that can be customized to serve decision-support needs. The work uses Hadoop as a tool to gather data from different sources in the healthcare environment and – by doing that – improves the quality of delivered information to governance boards. The work emphasizes that the utilization of big-data technologies has improved the level of acceptance in similar governance boards as to the reliability of imported data from the different data sources.

In [79] the concept is extended and scoped. The focus is on the idea of enhancing the data-mining approach to focus more on big-data implementation concepts. The essential concept is the implementation of the likelihood function for big-data models for healthcare data. The application of big-data approaches to data-mining techniques in healthcare results in an improvement of patient data quality. The work in [79] can serve as a proof of concept that can improve the quality of outcomes by utilizing data in symbolic formats, therefore extending the knowledge extraction process to benefit from big-data extensions.

The challenges of moving healthcare data to the big-data model are emphasized in [38]. The work addresses the data accessibility model that would require improvement when healthcare data grows to a scale that can no longer be managed by conventional data-mining techniques. The authors explain the stages of data access, starting from hot data and reaching archived cold data. The models then can be adapted to include data that is to be consumed at several intervals, therefore moving away from the classical hot/cold states to a more flexible hierarchy that helps achieve better analytical performance.

With the introduction of big data into the fields of healthcare analytics and decision-support, new challenges started to arise that were not possible to solve using classical data-mining and analysis methods. As an example of such a challenge, [55] approaches the problem of data privacy when performing queries against healthcare datasets. The queries that are sent to the underlying data store are executed and then the result is sent to a special process application that filters the fields based on a predefined list, and then removes them from the result and sends them to the caller. This is emphasized as a challenge in large datasets where the size is around 14GB or more. [55] shows that while such problems have been addressed and solved in classical data analytics, they are later encountered and exaggerated when big-data models are applied.

Information gathering from different sources in healthcare environments is another challenge for clinical decision systems. The variety of hospital systems and the growing size of data to be processed gives rise to a challenge that is addressed in [94]. The work aims at centralizing the data collection process by engaging both clinicians and patients in consuming structured data from various data sources. While some data formats like HL7 exist in healthcare environments and can be easily consumed, other data formats are harder to consume not only because of structural limitations but also due to regulatory constraints that allow only certain aspects of data to be collected.

Such limitations are also highlighted in [31], along with an effort to properly define the relationship between big-data and modern medical disciplines that generate large volumes of both structured and unstructured data. More often than other fields of research, medicine offers a challenge for data collection and organization, leading to quality decision-support. On the larger scale, it can also be argued that big-data models can be of better use than classical data-mining and knowledge-discovery techniques due to the multiplicity of data sources and the variety of structures, and due to the wide range of data producers such as internet health data and social interactions. The work emphasizes the fact that big-data is a better choice for processing data and for knowledge discovery, growing from a need to encompass large amounts of rapidly growing data into usable knowledge.

With so much data available in the healthcare sector, as well as in other sectors, questions are raised as to how much data should be gathered for analytical purposes, and once gathered, how should that data be maintained so that it would accurately describe the desired outcome. In [65] these two questions are explicitly stated with a comparative approach relating classical methods to modern big-data approaches. The authors stress the fact that as data-gathering devices increasingly become cheaper and more abundantly used, more data is to be expected that needs to be analysed. The big-data approach is to collect as much as possible of this data – if not all of it – into so-called data lakes. This approach requires that data-collection tools and methodologies exist that are proven and maintainable. The work stresses the fact that once data is properly collected in a big-data approach, then these data lakes can be accessed by analytical tools regardless of their original sources. The work does not emphasize the challenges with standardization and coding systems unification, but rather stresses more on how data is collected efficiently, through a model that they refer to as N1⁄4all.

Another important point regarding big-data analytics is explored in [90], and the question of how the data-search accuracy changes over a growing data-set are asked. The idea is explored using an example of simulating a data-search algorithm over a growing data size. The accuracy is measured at different data-sizes and is compared to the original data size. The work shows that as data grows to a certain scale where it qualifies to be labelled as “big”, then the search accuracy also increases until it reaches a saturation point. The reason behind this is that the size of the analysed data increases to a point where it exposes all the hidden or missing data-parts that reduce the search accuracy at low data-sizes. The authors experiment with large volumes of data on commercially affordable hardware that is cloud-based and back their findings with similar approaches taken by other researchers. The work in [90] is especially important to consider when gathering large volumes of data in an approach like [65] and [94] and taking into consideration the optimization in querying approaches in [31].

Given the above discussions about the size of the data, the different formats that healthcare data can be collected in, and the coverage of the data sources in terms of how much of the whole data should be consumed, [58] shows that it is also of interest to classify the data that is collected into two categories: the first category is granular data that exists in the collection set and is also the same as in the original set, and the second category is aggregate data that exists in the collection set but is not available or aggregated in the original set, but rather can be further disintegrated into transactional data. The work shows that when properly applied, aggregate collection methods can produce considerably equivalent outcomes compared to the collection of the entire data-sets. While this approach reduces the volume of data that is maintained in the decision-support systems, it requires a risk management process that ensures the quality of the aggregated data.

But in what order should data be prepared? [7] argues that when data is collected, it is accumulated in the form of time-based layers. The base layers serve as a guide in the structuring of later layers. In doing this, classical techniques such as Bayesian probability and frequentist approaches deliver a statistical problem-solving alternative to modern full-fletched big-data analysis. Previous knowledge of a certain cohort sample is clear guidance to future events. The mixture of classical probabilistic analysis with modern big data analysis is labelled as big-reasoning, an approach where the rigour of big-data is softened with accumulated knowledge from classical probabilistic reasoning. This ensures that groups of multi-disciplinary scholars such as clinicians, biologists and others can coordinate their contributions to a better understanding of privacy-constrained data volumes such as those found in medical health records.

Another characteristic of data is considered in [19]. Data collected to databases for analytical purposes should, according to the authors, have a special relationship regarding how it was acquired in the first place and then as to what it represents. The authors argue the structure of “Essentialization”, where data is collected based on the tight correlation among the underlying variables that created the data in the first place. This leads to data that can be meaningfully analysed to produce forensics-like investigations using big-data techniques. While the argument of correlation seems to clear the confusion regarding the structure and meaningful accumulation of different data sources, it fails to answer how health data should be analysed when most of the contributing data sources are heterogeneous by design.

An application of big-data analytics in the medical analysis is explained in [22]. The work focuses on oncological data as it is collected and populated in big-data analytical data warehouses and processed by MapReduce in a parallel query execution model. The main idea is the migration from an RDBMS-based analytical approach in oncology to a big-data one. The growth in volume and velocity of chemotherapeutical and radiotherapy data has set a stage to the application of big data, but the classical model for knowledge mining from RDBMS's is still in focus. The challenge is to keep the structure of data consistent as the warehouse is populated.

In the same regard, the data collection process is detailed in [75], providing a practical application of moving data from streams of high-volume healthcare databases, social networks, and medical devices to a usable data warehouse. While other works have focused on challenges such as data size, logical consistency and maintenance, [75] points to the more practical challenges that come with the data: data noise, varying data fluxes, and feature extractions. The proposed framework allows filtration of data at the source databases, systems and devices. The end result is a data warehouse with data that has undergone inconsistency checks and is valid for analysis. The data warehouse and framework will assume "clean" data, but it might also be the case that some of the data received are still in raw formats.

Usability of analytical results coming from big-data is also an important point to consider. Almost every work on big-data analytics in healthcare has had a focus on outputting decision-support results to clinicians and other major decision actors in a healthcare setting. [98] focuses on the needs of nurses, who are in direct interaction with the patient. The work argues that there is a need to move the knowledge acquired from big-data analysis to the other side of the healthcare spectrum, namely the nurse-to-patient relationship. This helps improve the quality of care by converting the knowledge acquired from the analysis of patient data to patient-centric knowledge that can improve the communication quality between nurses and patients.

Another complication in the big-data analytics approach is starting to emerge as more companies are moving their analysis engines to cloud-based hosts. The balance between low-cost scaled management of the underlying architecture and the growing risks of privacy and compromising sensitive patient records is approached in [86]. “Remote data auditing” methods are used in most enterprise hosted solutions, and the authors propose a technique for reducing the overhead of auditing newly uploaded data. The default was to redo the audit process all over again with every modification, but that creates an overhead that both the hosting company and the healthcare organization cannot afford. The approach presented here reduces this overhead by applying the audit process to the incremental change that is created in the data warehouse. [86] helps with promoting the “hosted big-data” approach for enterprise healthcare settings and could be included in proof-of-concept discussions with legacy solutions advocates to help them benefit from cloud solutions.

In addition to auditing challenges, there is another challenge that comes to play with the modernization of data mining techniques. While the classical methodology involved loading the data in offline mode or from daily backups, the growing demands for quasi-live analysis resulted in more frequent data pulls, where sometimes already loaded data is loaded again to ensure consistency of the analysis data warehouse. [46] suggest a solution that is based on a middleware component that manages the loading process by keeping track of loaded data-chunks and ensures completeness of loaded data and does not stress the online systems. It is clear from efforts such as [46] and [86] that the manageability of modern data mining techniques is a demanding topic. As the challenges grow with the modernization of big-data analytics, so does the research that leads to better understanding and operation of the new technology.

In summary, it has become important in the field of big data analysis that an informative approach is taken to move classical thinking to modern cloud-based, high-volume and real-time analysis. The challenges are not few and should be addressed with an incremental knowledge acquisition process that tackles all aspects of the problems at hand. The research has grown in recent years to cover these exact challenges, and the reviewed works in this section have been a selection of promising efforts to pave the way for better big-data healthcare experience. In addition to the presented work and arguments, it still needs to be proven that knowledge acquisition and production from raw and large-volume data is a scalable process, and not only a step beyond classical data scales. Such a concern is clearly addressed in [16], with a clear separation between data preparation and data analysis.

It is also worth noting that although big-data analysis has overshadowed the classical data mining skillsets accumulated over time, it is still possible to attack modern healthcare large-volume data analysis problems with a classical data-warehouse approach, not unlike the work presented in [101]. This requires a clear understanding of the existing data models that continue to change and get added to the analytical scope and required continuous adaptation of the data warehouse tables to the new models being included by the requirements. While this is a tedious task, it is not impossible to undertake it, so that in principle it cannot be stated that classical data warehouse techniques and methods are no longer applicable, but rather that their application comes at a larger cost than can be afforded by the rapidly growing healthcare data.

The choice between using one of the two data analysis models is, therefore, one that depends on the people preparing and consuming data and, on nature, and context of the analysed data.

3.3.4 Data Analytics

The focused clinical analytics domain does not benefit from all the advances in the global data analytics discipline. This is because most clinical research is focused on deriving patient-specific data intelligence from the large volume of available databases. Involving more functional characteristics in the healthcare domain to perform further analysis is not always possible. This is because clinicians and healthcare specialists are mostly concerned with satisfying KPI-scoped daily operations, which subsequently are based on the quality of service and the efficiency of resource management leading to better patient care. However, there has been a shift in recent years from pure patient-oriented clinical analytics to an extended scope where the patient is one of many players in the healthcare setting. In this regard, it is important to go through some research topics in the global discipline of data analytics and try to learn from the accumulated knowledge [13], [48] and [83]. A selection of research material has been gathered and is reviewed here.

An adaptation of the stratification-centric tranSMART system to the field of bionomics is discussed in [80]. The platform is improved through its API's to allow for structured phenotypic data to be included for analytical and predictive purposes. The work highlights the needs of some fields of biology regarding large numbers of properties.

As part of the data analytics literature review, a selection of publications was examined that has a close relationship to the topics of interest in this thesis. The major criteria for this selection were the role of business intelligence systems and frameworks in the healthcare informatics domain. Business intelligence in the healthcare domain could be considered as part of the clinical intelligence but could also be linked to data analytics as the solutions are usually derived from analytical models that were applied in the classical data analytics solutions.

Based on the definition of clinical intelligence from section 1.4, where it is closely defined with respect to business intelligence (BI), there is a need to examine how BI affects the healthcare decision-support processes by investigating the effectiveness of applying regular BI tools in the healthcare setting. As BI tools grow more effective and easier to use by non-technical users, it is interesting to see how this adoption has evolved in the healthcare setting and what challenges have faced its growth, and why BI tools in their generic definition cannot immediately be applied to the heterogeneous data sources that feed into the decision-support process in the healthcare informatics domain.

The effect of quality of the information system based on the usage of business intelligence (BI) tools in the healthcare domain are investigated in [28], where an evaluation approach is taken to determine the impact to quality of results in a hospital information system based on the proper application of the concepts implemented by a BI tool. The work does not specify a particular BI tool to evaluate, but rather focuses on the outcome of various attempts in different public hospitals. The results are studies to determine the relationship between the overall user satisfaction with the information system and the impact on the individual usage of the BI tools. In a general context, the work provides a study of several BI applications based on an existing HIS tool in some public hospitals.

In [88] the importance of BI tools is expressed in reference to the data warehousing techniques from a classical background. The work stresses the importance of exposing the various data sources into a single point of analysis which is the data warehouse for the BI tool. The data is analysed based on the needs of the organization and for the purpose of achieving a strategic and competitive advantage among other similar service providers. The historical data that is maintained in such a data warehouse is a source of understanding for the non-technical users who rely on the proper analysis of their different data sources to drive their business further.

The importance of BI systems is also discussed in [2] with a focus on small to medium-sized businesses. The work tries to identify the benefits that such a business would get out of using a BI tool after having partially or completely implemented an enterprise resource planning (ERP) system. The data that is generated from such an implementation over a long period of time is usually large and would also depend on how complete the adoption of the ERP system is. In many such businesses the systems are available for users but are not fully used due to the lack of management incentives or the limited number of human resources available. The work in [2] confirms the concept that good quality output from ERP systems is essential to leveraging BI tools in the later stages. The quality of data is a direct result of good training of the staff that use the ERP tool and a long record of good validation and verification practices in such organizations which will eventually delivery “clean” data to the BI tools. The better the structure of the organization the higher the quality and volume of data that goes into the data warehouses of the BI tools in use and would then serve as a building block for strategic reporting and decision support activities. The work in [2] suggests a deeper look into the ways to improve the quality of delivered data in terms of internal structures of the organization and not just in a general term.

The value of data going into the BI data warehouse for analytical purposes is found to grow as the size and speed of data generation increases. As soon as BI goes to real-time data the picture changes, as is explained in [78]. The work discusses the importance of big-data analytical concepts and how the process of cleaning the data from its original sources is prohibitively complex in a large analytical model. [78] defines a set of dimensions that are noted as important for accounting based BI tools, but that could also be considered of similar importance to other non-financial system BI tools. The dimensions defined are namely “accuracy, believability, objectivity, reputation, value-added, relevancy, timeliness, completeness, appropriate amount of data, interpretability, ease of understanding, representational consistency, concise representation, accessibility, and access security”. The most common and generally scoped dimensions are “representational consistency, accuracy, access security and relevancy”, and this is because in a different environment such as healthcare setting some patient records could be somewhat delayed without much impact to the consistency of the total medical file of the patient. The delay is assumed to be to a degree that does not compromise the consistency and structural integrity of the patient file. More importantly, the relevance of the records to previously recorded information is of critical value in the future BI analysis of these records. Decisions related to patient health and safety are based on both structure and completeness of the records, and also on their correlation with other patients in the same analytical context.

BI quality and the impact it has on the overall usage and utilization of the ERP systems is further explored in [66] with a stress on how BI tools usage and performance measurement outcomes are directly related and how BI usage leads to enhanced organizational performance improvements. The top management usage of the delivered BI information is directly affected by the diagnostic performance measurements that are loaded and analysed by the underlying warehouse and data analysis systems. The ERP systems would also deliver another type of performance measurement to the BI data warehouse. [66] serves as a guide to how performance measurements lead to better BI usage and utilization by top management personnel. This approach is based on delivering accurate and timely performance measurements in both interactive and diagnostic measurements and then consuming them in the BI tool through modern analytical methods.

In [70] the link between BI quality of information and the improved usage of BI tools for decision making in organizations. The focus of the work is to emphasise the role of the "information culture" when derived from the general culture of the organization, in particular with the body of users and information specialists and consumers who understand and benefit from the organizationally structured information pools to improve decision-making capabilities. The work looks at how researchers can define a company's information culture profile by looking at several aspects of the information used in a company, particularly where the aspects of the integrity of information, the formal specification of this information and the control, sharing and transparency of this information are properly put in place. The role of information quality is found to be very influential and affects how the decision-making users are motivated to using the underlying BI tools. The opposite is not true, in that the BI information usage and utilization do not feedback into the organization by improving the quality of information delivered to the BI tools. This lack of feedback is key to understanding why BI tools are not capable of influencing the way information is created in the organization. The long-term effects of BI analysis are not put in the scope of the work in [70], and for that matter, the lack of feedback that is identified in the work is limited to the short-term improvement in the organizational data quality. The concept of information sharing is key to understanding how improvements in data quality result in better utilization of BI results and outcomes. The work in [70] marks information-sharing as a key to giving stakeholders a way around the ERP data which is not already in the BI tool. This is important to note because not all data that is in the ERP system and that is of value to the decision-making process are already loaded in BI tools. The limitation of BI tools to the data that is provided to it or loaded into it by the technical personnel who prepare those tools for use means that stakeholders have access to the outcomes of the BI analysis that is delivered by the technical admins.

In [11] the need for a continuous evaluation of BI tools has to lead the authors to create their own BI system which they called BI4BI. The evaluation of a BI system is considered essential in the continuous decision-making process in any organization, and the work finds that the existing literature shows two major limitations in regard to how BI systems are currently being evaluated. The first limitation is the lack of clarity about the possible evaluation criteria when it comes to comparing and evaluating BI tools. This limitation is identified as crucial since it prevents stakeholders and decision-makers from performing a quantitative evaluation of the many BI tools that are available, and as such prevents the proper selection of a BI tool to perform the desired decision-support activity. The second limitation is identified by the authors as related to the proposed evaluation system that is in focus. This limitation means that any evaluation system that would be developed based on the existing literature and following the reviewed methods of evaluating a BI system would encounter problems and would generate incorrect evaluations. The authors in [11] attempted to create their own evaluation BI tool that would be used by decision-makers to select the most convenient BI system to use. This tool feeds in data about the various BI tools that are accessible and creates a set of reports that the stakeholders can use to guide them in their decisions. As the tool itself is also a BI tool, the authors attempt to evaluate it using their evaluation techniques. The authors find that tool is well suited for delivering accurate reports to the stakeholders and is found to be a useful reporting system for stakeholders of different usage capacities. The BI4BI tool uses a data warehouse to populate the information that it collects from pre-defined data sources about existing BI tools and uses a set of evaluation criteria that is then executed against the incoming data to present the output in the form of reports.

In [53] advanced and current techniques such as machine learning, deep learning, neural networks are examined and analysed in terms of their usage in modern application scenarios such as financial market predictions, and the impact they have over the classical approaches involving large data analysis. The work in [53] focuses on how machine learning can benefit business applications through improving predictability based on learning patterns in the data that are difficult for humans to detect. The work classifies machine learning into three categories, based on the existing understanding of the way machine learning can be used and adopted. The first category is supervised learning, where the system or tool is subjected to a structured set of data that is then processed and analysed. The outcome is then compared to the pre-known correct value and the system is informed of how far the prediction was from the actual result. The second class or category of machine learning systems is unsupervised learning, where the system is not trained in advance but uses a set of proven techniques that work in similar situations and applies those techniques to determine the prediction. The result is then compared to the actual outcome of the prediction and is then fine-tuned for future application. The third class is re-enforcement learning where the system is trained with a data set and the output is corrected after exposing the system to different datasets from real scenarios and forcing the system to change the weights of its internal neural network layers based on the forced adjustment by the professionals or by the distance from a successful result. While machine learning techniques have worked well in problems that are well defined and understood, they tend to fail to provide a satisfactory output when the problem is based on a skewed data set that is not similar to the previous training sets. The application to financial markets has such problems which tend to make the predictions of machine-learning systems only useful when the encountered data flux is comparable to previously encountered patterns.

A BI tool is presented in [40] which was developed for a hospital that required access to aggregated data from different data sources in its systems and from outside data sources as well. The BI system uses the three-layer architecture that is usually adopted by most BI and data analysis systems. First, the data is collected from data-sources in the different systems and is loaded into a data warehouse where it is later analysed by pre-built algorithms based on the requirements of the decision-makers. The data analysed is then presented in the form of visual reports. The system is pre-configured with the processing algorithms that have been developed by the technical users in the hospital and has in part replaced the classical way of generating and processing those results. The system in the work replaced the method of providing the entire set of data to the management and then having them process this data in their own tools and ways, which would usually lead to incorrect results and unmanageable decision-making processes. The tool in [40] is applying data analysis algorithms based on the business requirements by loading data from different data sources and storing them in a data warehouse for future analysis and processing. The BI system presented is focused on collecting data from the healthcare environment which is mostly fragmented, complex in its data structure, and that is patient-centric in such a way that its incorrect interpretation or even falsely acting against its improper aggregation may undermine patients' health and safety. The work is part of an attempt by the hospital for which it is developed to improve the visibility of the top management and allowing them to continuously observe how the implemented management indicators are changing. The set of indicators may vary from one hospital to another, but it does show that it is possible to reach a credible stakeholder decision-support system within a healthcare setting using classical data warehousing and data analysis techniques and concepts.

In [81] the relationship between the quality of the medical decisions and the usage of the BI tools is stressed, with material evidence through the statistical collection of data from the usage of several BI tools. As shown in other similar works that have been reviewed in this section, the usage of the BI tools was shown not to have a direct effect or influence of the quality of the underlying medical information that it is consuming. The relationship is only in one direction where the quality of the information promotes BI tools usage and leads to higher levels of user satisfaction due to the correctness of the results and the reduction of data cleaning that is required during the loading process.

[69] offers a high-level approach to the business problem of converting process-based queries to technical database-level queries that would then be executed against the underlying data in the data warehouse and produce outputs the could be converted to process output. These application scenarios are approached from a process perspective and not from a technical one. The work describes how a process owner would approach a business problem by thinking in terms of business models and how that process would feed back into the main workflow of the business, producing outputs that benefit the overall business model. The work avoids a technical approach which would then lead to another data warehouse that is crowded with technical queries that need a lot of time to process and generate little benefit to the business process that initiated them. [69] discusses process repositories and how they can be described using a non-procedural language which could be used by the business non-technical staff to create their processes. The challenge in [69] is to maintain a top-level approach by preventing a technical description of the business process, but at the same time the work reaches a stage where it needs to tackle data-level topics and consume the operations that are most usually needed such as create-read-update-delete (CRUD) operations.

While [69] offered a non-technical alternative to data warehouse analytics in a context that keeps the business owners engaged with a non-process data analytical model, the work in [73] puts data ahead of the process by focusing on a particular domain where data flow is large and is to be subjected to continuous analysis by market experts. The model is suited for an electricity company that wants to use business intelligence to improves its current processes for financial and operational purposes. The work in [73] designs and implements a self-service model where users create their own analytical units and run them against the large data repository in the data warehouse. [73] does this through an approach that is not based on the global best-practices of data warehousing and data analytics. In doing so, the work restricts its expansion to other business models that might also need such a self-service architecture. The system designed is linked directly to the incoming data flow and thus diverts from the recommended approach in data analytics, and in doing so the system itself would allow the non-technical users to analyse the incoming data flow but with the downside that they can do so using a small set of analytical methods that would not impact the overall enterprise system's performance.

[57] attempts to bridge the gaps that are usually found in implementations such as the ones described in [69] and [73]. The domain is also an electricity company that has similar requirements as those in [73], but the work in [57] tries to follow the best practices of building a data-warehouse business-intelligence system by loading the data from different data sources and importing the different structures into a single point of analysis which is continuously updated. The work is focused on a specific domain and is functionally designed to suit a single model of operations. Nevertheless, it could still be considered as an attempt to create a data warehouse for business intelligence based on the operational performance indicators that are suggested by the business process. Although

the work in [57] focuses on a specific application domain which would be similar to a healthcare setting in that the data sources are numerous, and the flow of data is continuous, it is unclear whether such an approach would produce good results in a healthcare setting where the data models of the incoming data fluxes are not only different but sometimes contradictory. The approach, however, is well structured and offers a basis that can be expanded to include more generic data models.

The evaluation of a BI system is always considered a key part of the overall adoption of a BI solution in a healthcare environment. [29] discusses maturity models which are currently adopted in the healthcare setting to evaluate the degree to which a BI system is matured, and it does that by comparing the desired maturity level of different dimensions under investigation against the actual maturity level of the BI system. The key tool to measuring this maturity level is a questionnaire approach where the BI users and stakeholders are asked to rank their experience and the current BI tool under use for the many dimensions that are collected from the literature. The answers are then scaled to reach a maturity level of the BI system. The approach gives a way of measuring a BI system's ability to deliver what it was adopted for and to get feedback from the key users about the progress of the BI system against previous measures. This measurement method helps to improve the BI system by shifting the focus of the architects and the system designers towards the most useful dimensions that serve the overall quality of their tool. It also serves as an evaluation methodology for evaluating BI implementations in a healthcare environment.

3.3.5 Deep-learning

Deep learning has emerged in the recent years as a powerful part of artificial intelligence (AI) that attempted to solve several problems with classical AI and the continuous dependency on humans to calibrate and adjust neural networks as they are applied to different problem types. An advantage of deep-learning is the possibility of moving learned knowledge from one part of the neural network application to another, in a process similar to how humans propagate knowledge across generations. [96] explains this technique through the automatic learning of feature layers in the neural network by applying a generalization learning procedure. The technique helps to reduce the neural network dimensionality by flattening out the number of inputs through an autoencoder method and a local denoising process. The result of this enhancement and direct implication of deep learning is shown in the application in domains such as two-dimensional space image generation and big data analytics, where volumes of data are efficiently analysed to decide which data sets are to be used for analysis. The technique also can be combined with electronic health records analysis of patients requiring palliative care.

Another application of deep-learning is presented in [105], where the concept of automatically adapting features in traditional neural networks is examined in relation to manual feature changes. The impact is studied in several application domains, and the work examines some multi-domain statistical features such as time-domain features, frequency-domain features and a combination of both. In another application, the work examines machine health monitoring tools, where the adjustment of features in their neural networks is both manually and automatically performed, and the outcome is compared where a mixture of both is used. The results in every application of the mixed feature adaptation show that a combination of manual and automated feature changes using deep learning shows better performance results.

[104] shows a more focused approach to applying deep learning to the healthcare domain. In the work, the authors use deep representation learning as an improvement to time-series comprehension of medical records through deeper analysis of text patterns. The work focuses on improving the machine learning capabilities of medical text in the time-variant healthcare records by identifying repeating text patterns and using them to improve the features of the machine learning tools. The outcome is a highly trained vector representation of the electronic health records with an ability to detect linguistic regularities and matches in language patterns. This application of deep learning to the medical domain shows that advances in deep learning can also be incorporated into the healthcare setting, either through improving the analysis of large text volumes in the medical records or by fine-tuning the existing devices and tools used to monitor patients and gather patient and clinical data [17].

3.4 Identified limitations in the reviewed literature

This chapter reviewed several works that contributed to healthcare-related disciplines and surveyed the most recent approaches in modern clinical and medical informatics. The reviewed literature was classified into multiple categories and was further investigated to construct a general idea of the current status of clinical decision-support in conjunction with domain-relevant methods and technologies. In this section, a top-view tabular reconstruction of the examined approaches is presented where the limitations in each approach and application procedure are clearly shown and highlighted. The result is a helpful tabular reference for the coming chapters where a need for a new approach to solving the problems identified in this and previous chapters is deduced. The limitations tabulation is a high-level construction and assesses qualitatively the limitations, showing the relevance of these limitations to the aim of this thesis.

Domain	Limitations	Relevance to Aim
Clinical Intelligence	Current approaches are limited to solving problems that are previously identified. Adaptation of new problem solution techniques is limited.	Directly Related
Clinical Analysis	As clinical data grows, classical analysis techniques no longer fit the need. Data models change and increase with the growth of healthcare scopes.	Directly Related
Big Data	Acceleration in healthcare data generation from different sources challenge the existing data warehousing technologies.	Indirectly Related
Data Analytics	Developing analytical algorithms can solve previously identified problems, but as new problems and questions come up, new ways of getting the answers need to be applied. This limits existing tools.	Indirectly Related
Deep Learning	Deep learning is increasingly being applied to improve and assist AI modules in healthcare such as devices and text pattern detection. This indirectly results in new structured data sets from unstructured data, bringing more data into the analytical domain.	Indirectly Related

Table 3.2: Limitations from reviewed literature in relation to the thesis aim

3.5 Summary, discussion and conclusion

The growth and diversity of the different disciplines of data analytics in various domains, particularly in the healthcare domain, gives a clear indicator of the direction in which the adoption and utilization of such technologies are pushing forward the decision-support demands and requirements and are leading to many new approaches within the same discipline to create the best implementation and benefit stakeholders the most. In this chapter, a review of the published literature was performed where the various domains of data analytics and its modern big data analytics subjects were explored with a focus on their direct relationship with the healthcare environment. The selection methodology for the literature review from Chapter 2 was applied and a wide range of published work was presented, discussed and objectively compared to similar works in the same subject matter. The outcome of this chapter is a good understanding of how other researchers and organizations have approached the problems of decision-support and data analytics in a heterogeneous environment such as the healthcare setting.

The variety of published approaches and the different domains that they were utilized in makes it difficult to shortlist the most efficient and successful approaches. However, the literature review in this chapter has shown that regardless of the domain and problem to be solved, the more an approach is structured against existing and proven methods of data analytics the more it is likely that this approach leads to a durable and reliable set of results and would later be extendible as the technology changes and the ways the data is analysed is improved.

Chapter 4

The Need for a Novel Approach

4.1 Introduction

The literature reviewed in the previous chapter spans the disciplines of clinical intelligence, clinical analysis, big data and data analytics. The limitations identified are rooted in the phases of data loading and transformation, and in the data analysis and improvement of results accuracy through further involvement of data consumers. The adaptation of data analytics and big-data models to healthcare analytics dictates that the various data sources that produce data streams be filtered not only for noise reduction, but also for data consistency within the analytical data warehouse. During the analytical processing phase, ambiguities in the imported data sets might result in false predictions, not because the data is incorrect but rather because correct data is incorrectly correlated. Furthermore, optimization of implemented analytical algorithms that deliver decision-support to data consumers can only be achieved through active involvement of data consumers in a way that is both user-friendly and technically effective. While each of the reviewed works had its own objectives, the general context of clinical analytics should be such that a continuous and self-adapting process can be implemented to deliver decision-support that is based on properly filtered and audited data with a technically-effective and user-friendly feedback mechanism that allows data consumers to review outcomes, interact with the analytical system to adjust parameters in an effort that future decision-support suggestions are less erroneous and contribute to better and more effective patient treatment.

4.2 Current Approaches from the Literature Review

The approaches reviewed in Chapter 3 give a clear indication of how the existing research has attempted to solve the problem of providing decision support and analysing clinical information in a scope that serves the data consumer the best. The works reviewed could be further filtered to focus most on approaches that delivery solutions that could be improved or that might have a limitation in terms of the desired results by the data consumers. To further clarify and focus on the limitations and gaps that exist in the current approaches, it is important to highlight the approaches that are of most relevance to the scope of this thesis. In contrast to the review methodology which was adopted in Chapter 3, this chapter focuses on relating the works that are most relevant to this thesis's aim and comparing them further with a detailed emphasis on the identified gaps and limitations, in particular towards their application to the dynamic nature of healthcare decision-support requirements. The approaches that were found most relevant to this thesis's aim are [85], [64], [6], [52], [9], and [12]. The emphasis is on each one's application to the problem of dynamic execution of an enquiry by a data consumer, using a comparative analysis in regard to the approach in each work, its applicability to the research problem, and the adaptations that are needed to make each approach useful in answering the research question.

4.2.1 The Clinical Data Intelligence Project

In [85], the foundation of the approach is probabilistic, in that it attempts to employ machine learning techniques to predict a data model based on a clinical database. This approach yields results that need to be further verified by the process and might not carry enough credibility to proceed with a certain clinical decision. Any probabilistic approach in data modelling requires a continuous correction algorithm that matches findings to the actual knowledge. This approach can be used to deduce a future occurrence or a potential problem escalation, for example in treatment plan analysis, but could also lead to wrong estimations when it comes to a value-based decision-support activity. In the sense of a predictive data model, [85] does prove that some problems in decision support that are fuzzy in nature – such as treatment plan failure – could be solved using a probabilistic model, but in many cases the accurate prediction, even for a short future time, is better than a fuzzy long-term prediction.

4.2.2 Data-Driven Exploration of Care Plans for Patients

The work in [85] could be compared to the approach in [64], where treatment plan history is matched for similar data sets such as symptoms, diagnosis or lab results, leading to a classification of the successful and the failed treatment plans. This classification approach is accurate and is based on solid data that comes from the existing patient data model. However, the information that can be reported from such a classification model is limited in terms of usability. The model cannot generate new information, but rather classify existing information to a certain degree of accuracy. The model requires input from the existing data pools and is thus backed up by a correct database of patient information. Classification models do require a pre-defined classification function to process the quantitative aspect of a set of similar data – treatment plans in the case of [64] – and perform the classification either by separation of the data into two groups (an acceptance/rejection algorithm) or an ordering algorithm where the weight of a subset is determined by the output of the applied function of inputs. The classification model in [64] is not temporal and would therefore need to be adjusted as the correctness factors change over time. For example, if treatment plan using antibiotics is effective today, it could be later found that these drugs have side effects that were not know at the time of development. In such a case, the algorithm would need to be adjusted to increase or decrease the weights of certain decision factors. The categories of input data are also pre-defined, so adding a new category of data – such as patient weight – would need to be adapted into the algorithm and developed again.

4.2.3 Knowledge management-enabled health care management systems

The approach in [9] is more relevant to a dynamic data analytical model in a healthcare setting. The work focuses on the correct classification of the data to be analysed before it is loaded into the data warehouse, taking into consideration the variation in data contents for different data sources. Structured data requires little or no model changes to be loaded, but non-structured data needs to be subjected to further classification and changes applied to their data model in order to bring them to a structured state where the data can be consumed. The system in [9] focuses on knowledge aspects related to the data in a healthcare setting, emphasizing the four stages of data transformation to knowledge, which are first the creation of knowledge from data, adding a structure to this knowledge, disseminating the knowledge and applying it to a particular situation or problem. To process the data into knowledge, the system constructs the meta-knowledge information from the host sources and then applies this meta-knowledge to the imported data. The knowledge warehouse is therefore populated with knowledge which is referenced in a meta-knowledge structure and is also clean and relational. Once the meta-knowledge and relational knowledge structure is established, the system in [9] allowed users to query and retrieve information about the data from multiple data systems. The methods by which the knowledge is queried, and the results are presented is not clearly stated, nor is it explained how the users interact with the system in a seamless and error-free manner. The work serves as a good description of methodology of knowledge management and analytical warehousing but is lacking the details that explain the use-cases and how the actual problems are solved using such an approach.

4.2.4 Framework for design and evaluation of complex interventions to improve health

Another work that targets the idea of enhanced end-user involvement in the decision support process is presented in [12], where the patient and the physician interact together in a classification-based decision system which suggests a treatment based on previous same-patient and multi-patient experiences. The shared decision model presented in this work aims at putting the decision-making activity in the control of the end-users, and in doing so brings adjusts the analytical algorithms that are running underneath the system layout by feeding back the end-user expectations and results into the algorithm parameters. Although this shared decision model is easy to utilize since it is presented in the form of a check-list where users “select” options that match their expectations, the system does not clearly explain how a new study could be introduced with a new set of questions and answers. The static nature of the system architecture makes it difficult to perform adjustments to existing decision algorithms but is – however - developed in a manner that makes it easy to use and adopt seamlessly in a clinical setup.

4.3 Current Limitations

There is a growing interest in supporting decisions for patient-facing actors such as doctors, nurses and other care givers, and for delivering information that is easy to consume for study-oriented processes such as third-party negotiations and drug management with pharmaceutical companies. The content delivered to outside partners is also subject to increasing research as to how it can be improved in terms of speed, data quality and automation. It should be noted that external as well as internal actors in the healthcare domain increasingly improve their functionality, because they are singular entities that consume well-defined data that can be easily consumed in vertical applications. Pharmaceutical companies are a good example, where they continuously invest in improving their automated procedures that give them the competitive advantage in dealing with hospital demands. Such companies know that if they can ensure proper and consistent data delivery to their customers then they can increase their profitability and improve their market share. Another example is third-party payers. These companies not only adhere to strict standards when it comes to data consumption and production, they stipulate their data standards and engage their customers – the healthcare providers – in fierce debates to ensure that they receive their patient's financial and medical data in the structures that they can easily consume.

The number of entity types engaging the healthcare providers are also growing. While classically there were a few external and internal "consumers" of healthcare data, more players have started to offer their services in new fields that were previously self-managed. Healthcare data now is consumed – in detail or in aggregations – by companies that offer consultancy services, quality standards, patient safety, ancillary and supporting services. The structure of a healthcare organization itself is growing, and specialized centres have turned to technology and high-end solutions to guarantee accuracy and reliability of patient care plans. The more data is produced the more the demand that it be presented, shared, or aggregated.

To serve as a summary of the limitation comparison in this section, the following table is presented. The table shows each of the implementations discussed in this section, along with three indicators that are evaluated based on the gaps found during the literature review. The first indicator shows the “dynamicity” of the implementation, in terms of how much relative effort is required by the users to achieve the results needed. The second indicator is the “flexibility” of the implementation, in terms of how easy it is to change it to adapt to a new set of user requirements and analytical needs. The third indicator is the “rigidity” of the implementation, in terms of how self-contained the data model is and how relational the data is maintained and to what extend the implementation can be used with a consistent and stable performance. In the tabular comparison presented below, each implementation will be referred to by the literature reference that it was presented in. As the indicator value is difficult to quantitatively measure, it will be evaluated on a scale of 5, with the following values: (1-Very Low, 2-Low, 3-Medium, 4-High, 5-Very High).

Implementation/Indicators	Dynamicity	Flexibility	Rigidity
[85]	2-Low	3-Medium	3-Medium
[64]	3-Medium	2-Low	3-Medium
[75]	2-Low	2-Low	4-High
[52]	2-Low	4-High	4-High
[9]	4-High	3-Medium	5-Very High
[12]	4-High	2-Low	5-Very High

Table 4.1: Summary of the most relevant implementations’ comparison

4.3.1 Data loading and maintenance

Analytical data is in constant motion. Data is produced in the transactions that the patient is involved in. The objective of every analytical engine reviewed in the previous section is to consume data in complex structures and generate usable analytical results that contribute to better decision support.

The improvement in technology has contributed to making this task easier, but on the other hand the same technological advanced resulted in larger and more complex automated information systems that produced more complex data structures with more inherent dependencies. The expectations of clinical decision makers had also increased. While a few years ago a simple diagnosis suggestion would have been considered a great achievement of artificial intelligence, anything shorter of a complex treatment plan that adapts to changes in patient conditions is considered a failure of modern clinical analysis. The state where data is loaded to the analytical engine and at which analytical processing begins is very critical. This stage was described in different contexts by the authors of [97], [89], [15], [49], [44] and [74] as part of their main analytical processes. The approaches are not always different, but some unique aspects of the various data extraction and transformation processes can be identified. [97] clearly loads the data into a statistical tool where it is analysed based on a predefined set of features that are pre-loaded into the tool.

[89] performs loading into a data warehouse with a clear focus on nomenclature that relates to the original drug-centric systems. The focus of [89] is to produce chemical analysis of drugs, and therefore the output is a name and composition suggestion. [15] uses text-processing with a matching algorithm to analyse input text and produce candidate values and mappings. The procedure is a single-input to a single-output

mechanism, which does not necessarily relay on pre-loaded or post-loaded information. This form of isolated data processing and analysis focuses on single-content sections of the entire data-set at a single point in time, therefore producing a time-specific result which is only of relevance to the time it was produced. [85] employs a large data warehouse for loading clinical information from healthcare systems, focusing on a probabilistic approach to generate results. Although the work does not focus much on the loading mechanism, the large scale of data loaded, and the variety of data formats implies a complex loading process. The data mapping is not stated, and therefore transitions from the original model to the model that that analytical engine processes is not clear, and therefore the variance between the original and the processed data is not determined. It is not clear how a change in the original data which requires propagation or addition to the data warehouse is carried out. This could be due to one of two possibilities; either the data is preloaded, analysed and then discarded, or the data is loaded every time a new analysis is required.

A one-time data loading process is then like approaches like [49], which targets a non-temporal small data-set with small attribute-set and therefore loading of this data is performed in a single operation per analysis. The analysis is performed using a statistical tool like the one used in [97], and the purpose of the analysis is to determine cases such as DISC diagnostic evaluation, leading to earlier diagnosis and treatment. A rather simplistic and statistical approach is also conducted in [44], where statistical and traditional spreadsheet tools are used to load and analyse “data chunks”, one piece at a time, in a similar approach as [97] and [49]. In such analytical techniques, the size of the analysed data-set and the complexity of the analysis algorithm must be such that they both can fit into the limitations of statistical and off-the-shelf tools. When any of those limitations is exceeded the entire analysis fails. Many small scale and single-study algorithms fall into this category and may utilize similar approaches. A larger healthcare-size data load is explored in [74], where the decision-support domain spans multiple disciplines within the medical spectrum. The work addresses Extraction-Transformation-Loading (ETL) as an intrinsic process to the healthcare analytics framework that is proposed [21]. The work even includes a reference to the variability of the ETL data selection mechanism, which is a defining property of healthcare data.

4.3.2 Involvement of data consumers

The most common purpose of a data warehouse and subsequently data analysis is to produce results that could be used to facilitate or speed-up processes such as decision-support, clinical treatments, or managerial actions. The final beneficiary of this analysis is the process owner behind this analytical effort. This could be a doctor in a clinic or a manager consuming results for better process optimization. This data consumer is usually a person employing the tool – the data analytics warehouse engine – to move from one state of a process to another. However, it is also important and very useful to the tool itself that the data consumer feedback to the tool with corrections or adjustments to parameters that could improve further analytical activities. This requires that the data consumer can measure the difference between the suggested decision and the correct one, consequently adjusting the tool parameters. Such an intervention means that the data consumer is an actor in the realm of the analytical tool. The degree of involvement that is found in the works reviewed previously is compared. Data consumers are found to be involved in the works presented in [82], [89], [64] and [75].

In [82] the data consumer is the radiologist receiving the images from the analytical engine. The radiologist compares the subtractions suggested with the ones detected using traditional radiology techniques retrieved from the PACS system. This comparison is repeated for several types of images and helps in improving the accuracy of the delivered suggestion. It is not clear how the radiologists' findings are matched with the image analysis performed on a technical level, and to what degree this affects the quality of the analysis performed. Nevertheless, the data consumer does have a role in the improvement of analytical results. A similar involvement is shown in [89], where the medical chemists are presented with drug structures that might help them link them to other parent structures, and in doing so they also identify those that form similar metabolites. The chemists also try to identify the routes through which those structures were formed. This link and originality assessment is done per presented result, and therefore the chemists contribute to the analysis quality in the long run. Like [82], [89] does not show a clear technical involvement plan for the data consumer. A more direct involvement is presented in [64], where the doctors receive the analysis in the form of similar treatment plans from previous patients. The doctor can interact with the system and identify the treatment plans that have the most accurate reasoning method. In doing so the doctor favours the reasoning paths based on his/her experience, and so improves the quality of the analysis algorithm. It is not clear how subsequent treatment suggestions are better than the older ones following this involvement, but it is a clear indication that the data consumer is actively involved in decision-support improvement.

The work in [75] is the one with most involvement from the side of the data consumers, and this is clearly approached from a design perspective in the last part of the multi-layered design. The proposed framework includes a mechanism where the data consumer monitors and measures the performance of the delivered results for accuracy and consistency. The work discusses the generalities of the big-data approaches and includes some improvements to the traditional processes. The degree of involvement is clearly states as a monitoring role. The data consumer has a feedback channel where inaccuracies are fed-back into the analytical engine. The framework does suggest an intermediate statistical approach, in that the data collected could be processed using off-the-shelf tools, but the output is delivered in a clear structure and format to the data consumer. The difference in approach between [75] and the other reviewed works is that [75] identifies the role and responsibilities of the data consumer, and in doing that includes the data consumer as an actor into the analytical evaluation model. The other works resorted to a single mention or statement of a use-case involving the data consumer.

4.3.3 Data quality and auditing

During all stages of the analytical systems and processes, the quality of the processed data is critical in the success of the delivered results. Data cleansing is a process that can be performed during all stages of the analytical model, starting at the data load and consolidation, moving to the processing and analysis phase, and finally in the reporting and visualization phase. The reviewed work involves multiple evidences of proper data quality practices, but the most notable works where data quality is particularly addressed are [92], [94], [31] and [75].

In [92] the idea is approached through reviewing governance-specific bodies that promote better information availability and data quality. The importance of this topic in modern big-data implementations is highlighted, but without a use-case for clinical analytics. The work serves as an introduction to how data-quality should be observed and in which cases it is particularly important. The approach in [94] is different, where data quality is addressed in the context of collaborative healthcare data contributions, with a practical approach of using the collective information of patient health data which has been approved by the patients themselves for usage in analytical models. This facilitates the creation of tools that help healthcare specialists efficiently utilize the large volumes of data and still ensure its quality. Both [92] and [94] offer a broad guideline into data quality best practices, without a practical implementation. Loading data from healthcare data sources should be done with care, as many data sources produce data that is neither reliable nor accurate.

This data is referred to as noisy data in [31], and the filtration of such data is a key factor to the overall quality of the analytical data warehouse. The authors state the problem of large volumes of data in petabyte sizes. The importance of data maintenance and continuous data quality auditing increases with large data volume intakes, and [31] focuses on the general best practices and usability designs for healthcare data mining and analysis. While the work does not depict any clear data quality maintenance model, it does stress the importance of following the best practices for big-data handling in healthcare settings.

On a practical level, the work in [75] stresses the importance of data quality in big-data healthcare implementations. The data loading process, which streams multiple data sources into the main analytical data warehouse, is equipped with data filters that allow for better noise reduction. The authors explain the fusing process where the velocity and types of data are of different scales depending on the source of this data. The mechanism for intelligent data fusion works to both integrate this data and improve its quality.

4.4 The need for a new approach

This thesis proposes a new and novel approach that is presented as a new system for clinical analytics. This approach is a new system that implements and adheres to the data analytics processes, and furthermore improves them by integrating real-time data quality checks in the loading stage and improving the analytical algorithms by engaging data consumers directly and easily into the decision-support process through feedback functions that allow them to optimize parameters.

The figure below shows the existing and current “common” clinical analytics model. This model is suggested by the revision of the state-of-the-art solutions presented in the reviewed literature and from current best-practices in modern clinical analytical and big-data disciplines.

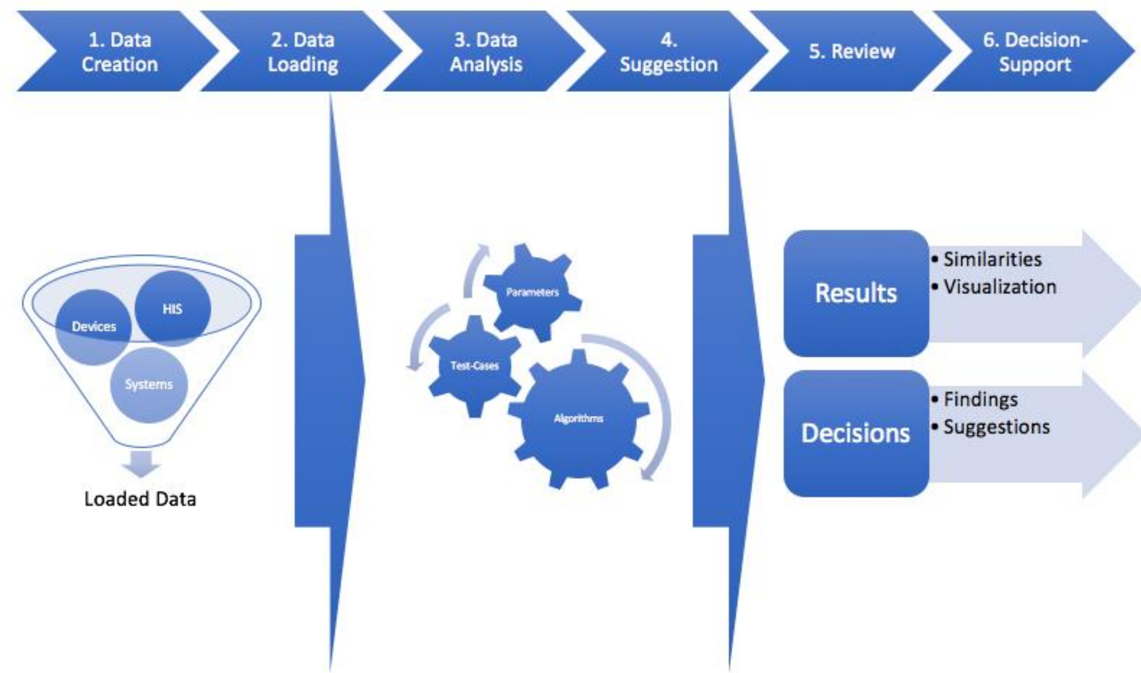

Figure 4.1: Current Clinical Analytics Process

The state-of-the-art systems for clinical analytics present the first phase in the above process as a consolidation phase, in that the overall objective is to collect as much data as possible to load into the data warehouse. The extraction, loading and transformation phase encapsulates data quality and auditing, and once the data passes into the data warehouse the information is already in a state that is ready to be processed. The limitation of this model is that insufficient data quality and auditing is done at this point, sometimes leading to inaccurate predictions and faulty results. The model also inherently assumes one or a limited set of algorithm implementations that target one or more use-cases from an application perspective. Most reviewed works adhere to a single analytical algorithm within the analytical phase. The final phase is the results presentation phase, where the decision-support is delivered to the data consumers. In the current model data consumers are recipients, in that they have limited or no possibility to influence the future behaviour of the analytical engines. As stated above, and since most algorithms implemented within the analytics phase are addressing a single use-case for decision-support, the implementation would not be flexible to allow feedback interactions from those data consumers. In that regard, the information flow is outward, initiating at the analytical phase and propagating in multiple presentation methods to the data consumers. The proposed method and system would be based on a modified and improved analytics process. In the figure below the conceptual process improvements are presented that will be implemented and presented in subsequent chapters.

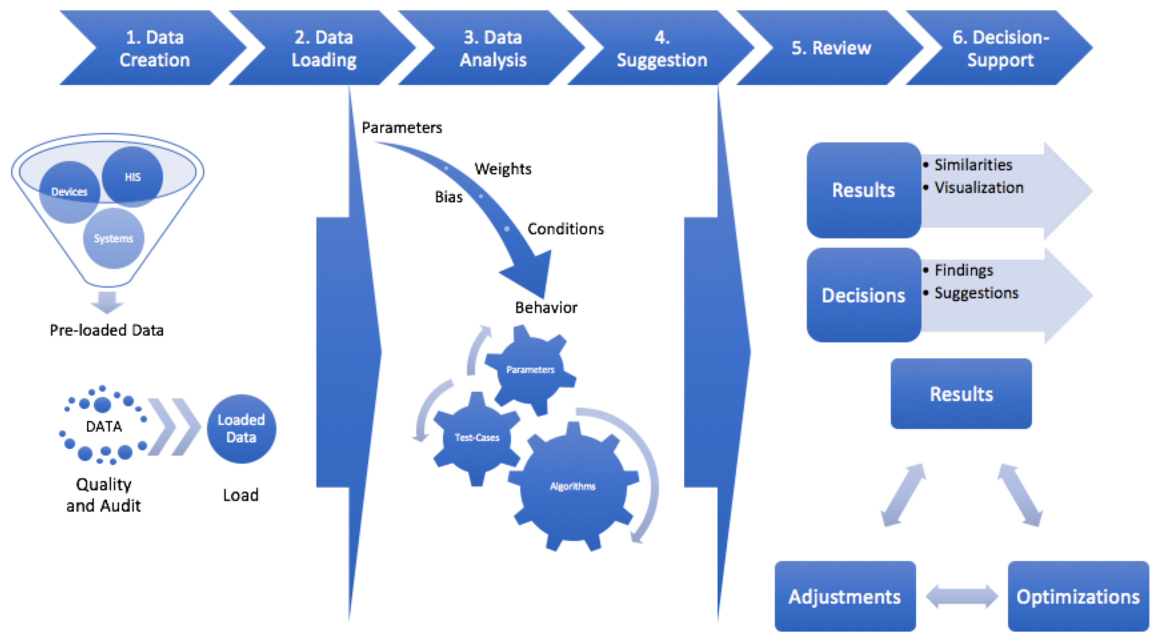

Figure 4.2: Proposed System Process

In the above-proposed process, the data analytical framework would apply a well-defined data quality and auditing mechanism that ensures that data is loaded in a consistent manner that minimizes noise from different data sources. Data loading is also audited through a sub-process that results in reliable data that could be changed from the source without impact to the subsequent phases on the analytics process. The second phase will include an improvement that introduces a parameters and scalar variables management mechanism which can then be used in the analytical engine to allow for multiple use-case and analytical algorithms. This empowers the functionality of the analytical engine and serves as an optimization model. The last phase is improved with a feedback and optimization sub-process that allows data consumers the possibility to influence future results and improve their accuracy, which was identified as a limitation of the current process and implementations.

In section 4.3 the limitations of the reviewed systems were presented and tabulated, with a scoping of each limitation in the major reviews where an analytical system was found. The three limitations that were identified have been addressed in the new approach that is proposed and are therefore resolved in ways that will be explained further.

The process of data loading and maintenance was found to be lacking key involvement from data consumers, which was also contributing to lower data quality in the analysed data pool. The proposed approach gives data consumers both the ability and governance to control the loading process by engaging them in the data definition and modelling phase. As there is a growing interest by data consumers to inspect and criticise the data models at work in their analysis tools, and with the flexibility of the new approach in terms of data model definitions, it becomes easy for data consumers to contribute to data quality improvement by controlling some data definitions. The proposed system does not give data consumers full authority in terms of changing the basic data structures, but it does present them with an option to shape their analysed data in the way they find most useful to their objectives.

The second limitation that was identified as the lack of involvement of data consumers in the analytical process itself. In the systems that were reviewed, algorithms applied to the loaded data were either created by technical personnel or aligned initially with data consumers and would never be changed. This prevents data consumers from contributing to the growth of the analytical models implemented in the system. As data consumers can now contribute to the structure of data that is loaded into the proposed system, it is thus a consequence of this that they are involved in defining the algorithms that are applied in the analytical phase.

The third limitation was identified in the overall quality of the analysed data, in terms of referential integrity, relevance to the analytic algorithms, and the produced output of the algorithms. The works reviewed do not include a process that maintains data quality, and the works that include basic attempts to perform data quality maintenance to the data warehouse systems were listed in Chapter 3. The proposed system approaches data quality based on the relevance of the existing data to the data models in the running algorithms. The data is loaded based on algorithms' requirements, which results in linking the algorithms with the loaded data sets. As algorithms might use existing data-sets, it is then possible to retire the data sets that are no longer used in analytical algorithms. The proposed system also includes a loading filtration mechanism that prevents "dirty" data from reaching the internal data structures.

The framework proposed is designed in such a way that it not only allows users to interact with a running system but also opens the doors for extending existing functionality by publishing Application Programming Interfaces (API's) that make it possible for building add-ons to the existing framework. The framework also allows clinicians and data-consumers to execute their own custom-built algorithms using modern user-friendly interfaces and a structured model for representing the loaded data structures. Clinicians are engaged from the loading stage onwards, giving them the opportunity to align their data ahead of the loading process. It also motivates them to experiment with their data and to re-use existing algorithms and extend them.

The framework design is multi-layered that limits the dependency between the system components, which is a key design decision that helps easily extend and customize the system for different process needs.

The framework is designed as two functional parts. The first part is the application that will capture the data from different sources through loading or polling mechanisms and will also allow for clinicians' interaction with the loaded data through the configuration of loading transformations that shape the resulting data in the data warehouse. The application will also host the algorithms that are created by the clinicians and other pre-loaded algorithms. The application will process the data and visualize it back to the users through client-side widgets that are easily integrated into many applications. The second part of the framework will be the integration components, which expose almost every data aspect of the application to other systems. This API-based interface capability enables application developers to load analytical results, structures and outcomes to other systems in real-time, a functionality that is currently not available with other similar systems. The application itself will utilize this API functionality to communicate data to and from the application server, and the API functionality is also flexible to enable different transport formats, including SOAP, JSON, and standard XML.

4.5 Summary, discussion and conclusion

In this section, the current approaches were highlighted based on the review of the literature from Chapter 3. The most relevant approaches with a similar analytical design were compared again and presented in a result-centric approach to reveal the similarities and differences among them. By doing so, this chapter established the existence of a gap in the current approaches which is mainly shown in the way the end-users interact with the decision support systems. The need for a dynamic meta-model approach is clearly explained in relevance to the existing static meta-model approaches, giving a clear view of what needs to be done in relation to closing this gap and designing a dynamic meta-model for clinical analytical problem-solving and decision-support needs. The clinical model described in this chapter will be later expanded and analysed in subsequent chapters, where a strict and systematic approach will be taken to define the software architecture of the intended system framework, followed by an evaluation based on the methodologies presented in Chapter 2.

Chapter 5

Analysis

5.1 Introduction

In the context of clinical practices, it is a clear requirement that clinicians can manage the data coming from various healthcare information systems and conduct an objective assessment of the patient's health status to benefit from previous experiences and similar cases to achieve an efficient treatment plan, minimizing side-effects and increasing patient comfort. A proper understanding of the patient's medical history and the incidents leading to the current medical condition is required, in addition to a series of diagnostic examinations such as laboratory and radiology exams to determine a proper clinical diagnosis for the patient.

A patient's encounter within a healthcare setting can be divided into several stages. If the patient visits a doctor's clinic and exhibits conditions that require further interventions, the doctor refers the patient for hospitalization. The patient may also be admitted through the emergency department following a triage, where it becomes obvious that hospitalization is required. At this stage, the patient is referred to as an inpatient.

It is a common practice at most hospitals that inpatients undergo regular exams to properly diagnose and assess the medical condition. These exams are usually executed by the admitting doctor's orders, and their results are either available prior to admission or immediately on the first day of admission. The objective of these diagnostic examinations is the appropriate selection of a treatment plan by the attending doctor which results in a series of medical orders to be executed by the hospital staff, starting with the nursing team. The patient will undergo further monitoring exams, sometimes in repetition of previously prescribed ones, to determine the efficiency of the treatment plan and observe any side effects. Vitals signs are regularly recorded to produce a clinical description of the bodily functions of the patient during the entire hospitalization period.

During most encounters, the doctor will be able to correctly diagnose the patient following the initial observations and exams. However, in some situations the initial treatment plan fails to deliver effective results. The patient might be allergic to the medications prescribed or might be resistive to certain antibiotics. In critical and complicated cases, the treatment plan might not be suitable due to other complicating factors such as physical inability to perform bodily functions or malfunctioning body organs.

While no two patients have the same exact medical situations, it is possible to associate symptoms and diagnoses with a group of patients. Clinicians often face the problem of associating the medical condition with the previously used treatment plans. This is never a one-to-one relationship, so in almost all cases a patient's treatment plan is an adaptation of previously – and successfully – executed plans. A treatment plan is not restricted to drug prescription and might involve other therapeutic measures.

Records of the patient's hospitalization and treatment are kept in the patient's medical records at the healthcare organization and appended to his/her health records. Healthcare clinician's usually conduct studies involving several patients with a common set of diagnoses, and such studies might be financed by insurance companies or research centres interested in the efficiency of certain medications or medical equipment. Such studies usually consume a great part of the clinician's time, requiring collection of data in certain formats to deliver the desired results. As was evident in the review of current literature, it is usually the case that clinicians use general-purpose tools such as spreadsheet processors and statistical packages. The clinician is required to compile outcomes from these tools and create appropriate reports.

Healthcare data is also consumed for other purposes such as education, financial cost-benefit studies, and organizational expansion/investment. Healthcare management will also require similar reports to estimate the efficiency of the working staff, and to highlight recurring problems in departments that require further interventions. Stakeholders might also require quarterly performance reports that show the overall productivity and profitability of their organization.

While most healthcare information systems provide reporting and statistical functionalities that combine transactions into consolidated views, it is usually the task of clinicians and information technicians to collect such reports and utilize them in delivering the end-result. While this might be a straight-forward task in small organizations, it quickly becomes forbiddingly time-consuming in larger organizations. In addition, the process of generating the reports and submitting them usually results in repetitive tasks to eventually refine the outcome based on adaptations to changing requirements. In this agile environment, it becomes essential to automate several data extraction and processing tasks, and when the information system does not easily provide such a functionality, it becomes the clinician's burden.

5.2 Case Study: a survey from the healthcare setting

In order to explain the current situation in the local healthcare environment, a case study was conducted where four hospitals were surveyed, and subjects were chosen to answer a questionnaire and the subjects were interviewed. The survey helped construct a statistical identification of the local situation. The case study was started by selecting groups of personnel in direct contact with the E.P.R. from a range of healthcare centres in the country. The selection criteria were based on the nature of interaction between the subject and the E.P.R. in terms of information retrieval, decision support and clinical findings/research. The subjects selected were unit/department managers, statisticians, decision support managers, doctors, executives in the financial, medical and business domains, information analysts and in some cases experienced E.P.R. personnel. Excluded were employees who perform more transactional updates than queries through reporting tools. Also excluded were personnel who control information through a single component and are not involved in data from other E.P.R. components, the reason for this being that most questions depend on the subjects' need to access information beyond the primary component that they use.

Electronic patient record systems present a challenge when patient records need to be queries across different components of the E.P.R. systems. With different components of the E.P.R. co-existing in the healthcare eco-system, it is usually the case that information about a patient needs to be retrieved for comparison and analysis. This information is distributed across different E.P.R. components or modules with which the patient interacted during the providing of the health services. When such information is needed, the components need to expose their internal structure or schema for extracting the data and executing the query. It is customary that E.P.R. components from different software vendors exist, and that these components hold information about the patients in their own storage engines.

The purpose of the case study is to assess the validity of criteria through a questionnaire targeting a sample of the population of all professionals in the healthcare local domain. With the responses that this thesis collects from the subjects, this thesis can decide on the validity of the gap that was identified in the literature review from Chapter 3. This thesis will therefore establish the correctness of the gap identified and use the questionnaire results will be analysed to prove the existence of the gap and the problems that exist due to this gap.

The case study gathered responses from professionals in the healthcare domain with varying skills and specialties, and generated questions that will help estimate through the sample of subjects the conclusions from the literature review of the current situation.

In this thesis, a questionnaire was designed to collect feedback from personnel involved in querying and data feeding from the Electronic patient record systems that co-exist in their environment. It asks for information about the relationships between the subject's queries and the domain of interest of his/her research or work. The number of subjects for the questionnaire and the scope of questions and desired results favoured an interview-administered survey. Subjects were interviewed in their work location and the questioned were explained when needed, and answers were directly recorded.

The design of the survey comes in alignment with the approach in [10], where a set of design and feasibility points are raised as guidance to implementing a case study survey. [10] points that in order to design a successful questionnaire survey, and to avoid confusions related to content of the questions and the overall goal of the survey, the questions should have a clear content, can be administered in a time-frame that allows to the personnel being surveyed to answer all questions but not disturb their normal daily work, and that the survey questions be decided in such a way that they avoid unnecessary questions that add no value to the survey aim, have a single quantifiable outcome for the users to choose from, and generate a measurable value that can be analysed from the results of the survey. Adding a few free-text questions is also a recommended survey design choice. As a result, [10] points that the surveyed personnel should be clearly notified of the questionnaire being undertaken, the time they need to perform it, and the desired outcome of the entire survey. Moreover, the survey should be formally introduced to the participants, and that they know before taking the survey that their identity will remain anonymous. This thesis followed these suggested points from [10] and maintained a professional engagement approach with the surveyed personnel.

The surveyed personnel were not all of the same academic or professional level, and the first section of the survey helps group the participants based on years of experience and educational and professional qualifications, leading to the identification of a focus group whose input in the pre-survey meetings helped propel the study further by highlighting the important aspects of the healthcare domain and the need for a similar work as the one performed in this thesis to extend the ability of decision-support and reduce the need for manual and effort-consuming activities.

While this choice of questionnaire methodology prolongs the duration of the questionnaire and answer collection, it reduces bias and ambiguity of analysed data and allows for a better understanding of the requirements and concerns of the subjects, as observed in [72] with a reference to an indirect approach that “Other research has taken” by “examining the link between problem diagnosis and specific response patterns (for example, missing data, or ‘seam bias’), on the assumption that higher or lower levels are more accurate". This approach, according to [72], is preferred to the error-predictive approach taken by some other questionnaire research.

The interview material will later serve as a basis for requirements and gap analysis. The choice of question formats was based on the desired method of analysis and fact-finding techniques. Open-ended questions were avoided, and Likert-scale questions were mostly used. Some dichotomous category questions were employed to classify subjects and provide an explanation of variances in answers to other questions. Due to the fact that the subjects interviewed were in senior and middle management positions, sorting and ordering questions were not used to avoid speculative answers. Both nominal and ordinal polytomous questions were avoided.

The choice was to reduce the number of questions and avoid negation questions. The comments of the interviewed subject were recorded for each question and every question was explained when needed, thus eliminating the need to detect the subject's random choice of answers through repetitive questions. The subjects were chosen from a larger population of candidates, and the choice was based on the effectiveness of the interviews and knowledge of the problem domain.

The questionnaire is shown below. The sections are labelled based on the grouping of the questions with a common objective. In the analysis to follow, questions within the same group/section are more likely to identify a fact than questions from different sections combined.

Querying the E.H.R. - Questionnaire

Thank you for taking the time to fill in the responses to each of the following questions. As part of our interview, I will take note of our discussions and will benefit from your experience and subject knowledge.

Section 1: Demographics				
Do you have an E.H.R. system at your workplace?				
	Yes		No	
Which E.H.R. system are you using?				
A	B	C	D	Other
What is the type of hospital that you work in?				
	Public		Private	
How many beds do you have at your hospital? ___________________				
Section 2: Involvement with the E.H.R.				
One or more of the E.H.R. systems is accessible to you at your workstation.				
Never	Rarely	Sometimes	Mostly	Always
Your E.H.R. system allows extracting data.				
Never	Rarely	Sometimes	Mostly	Always
You use third-party software to process data from the E.H.R. (Excel, SPSS, …)				
Never	Rarely	Sometimes	Mostly	Always
Section 3: Processing and Analysis				
The E.H.R. allows de-identification of medical data.				
Never	Rarely	Sometimes	Mostly	Always
You send data from the E.H.R. to parties outside your work premises.				

1

Figure 5.1: Case Study Questionnaire - Page 1

Never	Rarely	Sometimes	Mostly	Always
You contact others to get data from E.H.R. modules that you cannot access.				
Never	Rarely	Sometimes	Mostly	Always
Section 4: Collaboration and Decision-Support				
You require IT assistance to access data that the E.H.R. does not expose.				
Never	Rarely	Sometimes	Mostly	Always
Data that you export/process is stored locally on your computer.				
Never	Rarely	Sometimes	Mostly	Always
You collaborate with others who process different data from the E.H.R.				
Never	Rarely	Sometimes	Mostly	Always
Section 5: Security and Risk Assessment				
You use the E.H.R. for data extraction from outside your work premises				
Never	Rarely	Sometimes	Mostly	Always
You perform cross validation on data extracted from the E.H.R.				
Never	Rarely	Sometimes	Mostly	Always
E.H.R. data is filtered based on your access privileges.				
Never	Rarely	Sometimes	Mostly	Always
The E.H.R. only allows extraction of archived data (without further modification).				
Never	Rarely	Sometimes	Mostly	Always
You have unrestricted access to historical E.H.R. data.				
Never	Rarely	Sometimes	Mostly	Always
Section 6: Complexity of Analysis tasks				
You request the assistance of IT to organize extracted E.H.R. data				
Never	Rarely	Sometimes	Mostly	Always
The E.H.R. allows scheduling of automated extraction and delivery of data				

2

Figure 5.2: Case Study Questionnaire - Page 2

Never	Rarely	Sometimes	Mostly	Always
Audits are performed based on exported data from the E.H.R.				
Never	Rarely	Sometimes	Mostly	Always
The E.H.R. allows uploading and sharing your processed data.				
Never	Rarely	Sometimes	Mostly	Always
Your data processing methods are open to others in the organization				
Never	Rarely	Sometimes	Mostly	Always
Section 7: Identification (HCODE: _____)				
Your role at the organization is:				
Financial	Medical	Nursing	Clinical	Analytical
How many years have you been performing this role at this or other organizations:				
0 - 2	3 - 5	6-10	11-15	16+
What is your highest level of education?				
Bachelor	Master	Doctoral	M.D.	Other
What is your primary E.H.R. component/module?				
Financial	Clinical	Medical Records	Ancillary	Other
How much are you satisfied with the E.H.R. system that you are using?				
____				
____				
____				

3

Figure 5.3: Case Study Questionnaire - Page 3

The questionnaire was presented to the participants with an introduction letter that explains the purpose of the study and why this thesis is important to the fulfillment of an aim. The participants were guaranteed that their participation will remain anonymous and their identities will not be at any point disclosed. Below is a copy of this letter:

Monday, July 29, 2013

Dear Participant:

My name is Fadi Nammour and I am a PhD student at Kingston University - London. As part of my research involving the healthcare sector and its needs, I am examining the problems related to the *querying of the Electronic Health Record (E.H.R)*. The E.H.R. is used at hospitals and healthcare centers as a repository of patients' information during the several encounters that the patient has with the different departments/centers at one or many such points of health services. Professionals like yourself find themselves needing information from the E.H.R. and need to access this information on their own and from the comfort of their stations. They need to query this E.H.R. to find relationships, detect problems and perform various statistics and analysis on patients' data.

Because of your professional status and expertise, I am inviting you to participate in this research study by completing the attached survey.

The following questionnaire will require approximately 30 minutes to complete. In order to ensure that all information will remain confidential, please do not include your name. Copies of the project will be provided to Kingston University – London.

The information that you will provide, both in your replies to the questions or as part of our interview, will remain confidential and anonymous. This information will only serve as part of the research leading to my degree. I hereby assure that I will adhere to the ethical codes and moral commitments to ensure no leakage, disclosure, alteration or identification of this information, in part or in total, and that following the achievement of my purpose in my research, this information will not be available for any other use.

Thank you for taking the time to assist me in my educational endeavors. The data collected will provide useful information regarding the current status of the healthcare sector and the possibility of improving it.

Sincerely,

Fadi L. Nammour

Email: k1162086@kingston.ac.uk

Telephone: +9613283210

fadi nammour

Figure 5.4: Case Study Questionnaire Letter

The questionnaire was distributed to the participants following a short interview where the questions were explained when needed and the general healthcare situation was discussed.

5.3 Case Study Analysis

Ideas from a focus group of analytical staff in the healthcare organizations were put together to further identify the limitations and problems of the local situation. The group emphasized that data from current E.P.R. systems cannot be analysed without some sort of intervention from technical specialists, either in the preparation of data or in the execution of queries against the databases. The lack of a dynamic analytical framework results in the continuous dependency on these technical professionals, which is a delaying factor in managerial decision making. In many cases it was discussed that management often does not wish to involve technical staff in managerial decisions, thus leading to a conflict of interest between the efficiency of data analysis and the privacy of business intelligence. Organizations dealing with researchers, such as hospitals with specialized cancer centres or diagnostic facilities sometimes need to provide analysis from similar centres to improve the quality of care and performance of the organization. In such a situation, the availability of aggregated analytics is subject to requesting it from the technical departments across these organizations. This leads to problems since organizations use the privacy of patient information as a deterring excuse.

On the other hand, the focus groups stated that government agencies often require healthcare organizations to present aggregated data in the form of answers to specific questions (for example the total number of deliveries from a particular country in a period of time) and these usually can be answered directly by non-technical staff. However, this aggregated information is gathered in a central repository but not shared among the contributing organizations. There are technical and legislative barriers that prevent the secure and private exchange of aggregated healthcare information among researchers and doctors nationwide.

5.3.1 Case Study Hypotheses

The purpose of the case study is to assess the validity of criteria through a questionnaire targeting a sample of the population of all professionals in the healthcare local domain. With the responses that this thesis collects from the subjects, this thesis can decide on the validity of the following hypotheses that this thesis have compiled from observations in the field.

This thesis suspects that the local healthcare domain lacks a solid foundation for inferring knowledge from electronic patient record data. This thesis also suspects that most healthcare centres (hospitals, care centres, clinics, etc…) use health record systems that might be partially or totally electronic. It is also likely that there exists a data extraction and analysis process at these centres, and that professionals use data from the different electronic sources in the E.P.R. to support decisions by their senior management. This thesis will therefore establish our hypotheses and use the questionnaire results to prove them, or otherwise negate them. By doing this thesis will have established a picture of the local domain's involvement and level of utilization of E.P.R. data. This thesis will test the following hypotheses:

1. Clinical E.P.R. users share their processing methods more than Financial E.P.R. users.

2. Users who have been in the domain for more than 10 years need less IT assistance to access data from the E.P.R. system.
3. Doctors access the E.P.R. from outside their work premises more often than others.

This thesis conducts the case study to gather responses from professionals in the healthcare domain with varying skills and specialties. Based on our proposed hypotheses, this thesis will generate questions that will help us estimate through our sample of subjects the accuracy of our statements.

5.3.2 Statistical Analysis

The responses gathered in the survey would be analysed using the SPSS software tool. This thesis will first perform a descriptive statistic of the responses to identify questions that might result in misleading responses. The first step is to run the "Frequencies" descriptive statistics on the responses data set. This thesis will look for the skew in the distribution, the dispersion range, standard deviation, and standard error mean. These values will allow to understand how the answers are distributed and what questions lead have higher dispersion than the others. From the resulting analysis, this thesis will also divide the skew by its standard error, and questions with an absolute value of 1.9 or more will be considered as having a high skew and therefore might lead to misleading results.

The range of the responses will indicate what responses were never chosen. Typically, in a random sampling of subjects this thesis should expect the range to cover all the response values, but in this case this thesis will allow for ranges that are less than 5, because the questionnaire targets professionals in the domain and not every member of the healthcare sector.

The standard deviation will help identify questions that caused a wide variation in the responses or a similar response range. Responses with a very large standard deviation might result from questions that were not seriously considered by the subjects, and therefore need to be stressed and further explained during the interview. On the other hand, questions with a very small standard deviation might be biased and need to be restated in a way that does not favour a particular answer. In addition, this thesis will need to refrain from discussing such questions ahead of the subject’s answer, as this thesis might be inducing a bias to this thesis’s point of view.

Following the descriptive statistics, this thesis will look for correlations among the posted questions. This thesis will look for correlating questions within the same section of the questionnaire. Given that the questions are based on experiences of professionals in the same sector, and that this thesis’s purpose is to identify the weakness in the analytical processing of electronic patient records, this thesis will therefore assume that the questions are linearly correlated. Following the corrections that this thesis makes to ensure that the responses are normally distributed, this thesis can now look for the Pearson correlation coefficient in the responses. Questions in the same section of the questionnaire should have a correlation coefficient close to 1, indicating a positive relationship. When the coefficient is negative, this thesis will reverse the statement in which the question is asked to reduce negative relationships.

The questionnaire is not ordered. In that sense, this thesis can randomize the order of questions within the same section and expect similar responses. This thesis will therefore choose two-tailed tests of significance when looking for the correlating questions.

To further analyse the data from the case study, it was loaded into SPSS to deduce the relevance of the research question in the different sections of the questionnaire using the 5-grade Likert scale. The figure below shows the distribution of question answers against the individual participants:

3	Sum of Value	Q																					
4		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24
5	1	2	0	0	0	5	4	4	5	1	3	2	5	5	3	3	2	5	3	4	5	4	5	2	0
6	2	2	0	0	0	4	5	3	5	2	3	1	4	5	2	5	1	4	5	4	5	5	4	1	0
7	3	2	0	0	0	5	5	4	5	2	3	2	5	4	3	3	1	5	3	5	5	5	4	1	0
8	4	2	0	0	0	5	5	4	4	1	1	2	5	5	3	1	3	4	1	3	4	5	5	1	0
9	5	2	0	0	0	5	4	3	4	1	3	2	4	5	3	4	1	5	4	4	4	5	5	2	0
10	6	2	0	0	0	5	5	3	5	2	3	3	5	4	3	4	1	5	5	3	5	5	5	1	0
11	7	2	0	0	0	5	5	5	4	2	1	5	1	3	1	5	1	5	3	3	1	3	1	2	0
12	8	2	0	0	0	5	3	5	2	3	1	4	3	4	2	3	2	1	4	3	2	4	5	4	0
13	9	2	0	0	0	5	4	4	3	3	2	3	4	4	1	3	1	1	5	3	2	4	4	4	0
14	10	2	0	0	0	5	5	5	1	1	1	1	4	1	1	1	1	5	5	3	1	5	1	1	0
15	11	2	0	0	0	5	4	5	2	1	4	5	4	4	2	1	1	5	4	4	1	5	2	3	0
16	12	2	0	0	0	5	5	5	1	2	4	5	5	4	1	1	2	5	5	3	1	5	1	3	0
17	13	2	0	0	0	4	5	4	2	2	4	5	4	5	1	2	1	5	4	3	1	5	2	3	0
18	14	2	0	0	0	5	1	3	1	1	1	3	1	3	1	5	1	5	5	1	1	3	1	4	0
19	15	2	0	0	0	5	5	4	3	1	1	3	5	1	1	5	5	5	5	1	1	5	1	1	0
20	16	2	0	0	0	5	5	5	1	4	4	4	5	4	1	3	1	1	5	1	1	4	1	1	0
21	17	2	0	0	0	5	5	5	1	1	4	5	5	4	1	5	5	1	1	4	2	5	1	2	0
22	18	2	0	0	0	4	4	5	1	1	1	3	1	3	1	5	1	4	4	1	1	3	4	1	0
23	19	2	0	0	0	5	5	5	1	2	3	3	5	3	1	3	1	5	1	1	1	5	1	5	0
24	20	2	0	0	0	5	5	5	1	1	4	4	5	5	1	4	1	5	5	3	1	3	1	3	0
25	21	2	0	0	0	5	5	2	1	2	1	4	1	3	1	2	1	4	4	3	1	5	1	5	0
26	22	2	0	0	0	4	5	4	1	1	1	5	1	3	1	1	1	5	5	5	1	3	1	1	0
27	23	2	0	0	0	4	4	4	2	1	3	2	4	2	1	1	4	3	3	1	1	1	1	2	0
28	24	2	0	0	0	5	5	5	1	1	4	3	5	2	1	1	1	5	5	1	1	1	1	1	0
29	25	2	0	0	0	4	5	5	4	5	1	5	3	4	1	5	5	0	5	3	2	4	4	3	0

Figure 5.5: Case study answers by the participants

The were some questions which had a subjective answer, and where not included as part of the quantitative statistical analysis.

Having stated the research hypotheses earlier, the truth of these hypotheses is examined based on the responses collected from the subjects.

To test the hypotheses this thesis will use one-way ANOVA (Analysis of Variance) to determine if the means of the dependent variables are statistically dependent on the independent variable groups. Tukey post-hoc tests will reveal this significance among group members, while the descriptive statistics will provide means and standard deviation calculations for comparison purposes. This thesis will test for the significance level of 0.05.

The first hypothesis is that Clinical E.P.R. users share their processing methods more than Financial E.P.R. users. Clinical users are mostly nurse managers and physicians who attend to patients and are focused on the health problems that need to be solved. In this hypothesis this thesis is testing the openness of users to each other when performing analytical processing of the information available in the E.P.R. system. A clinical example is the sharing of the method to determine the relationship between WBC values in babies that are admitted with high fever, while a financial example can be the method by which the cost of non-chargeable items in a closed unit can be related to admissions with a specific diagnosis. These methods of analysis are the building blocks of the analytical structure of knowledge that drives decision-support. This thesis used question 27 as the independent variable or factor, and question 23 as the dependent variable.

Question 27 groups the subjects by asking them about their primary E.P.R. component that they use, while question 23 inquiries about the level of sharing of their processing methods. Based on the collected responses, there was a statistically significant difference between groups as determined by one-way ANOVA ($F(3,21) = 5.129$, $p = 0.008$). A

Tukey post-hoc test revealed that the openness of the processing methods was statistically significantly lower with financial users (1.43 ± 0.535, $p = 0.031$) compared to the clinical users (3.13 ± 1.246). There were no statistically significant differences between the medical records group ($p = 0.999$), but a very close to the significance level for the ancillary ($p=0.063$).

The second hypothesis is that users who have been in the domain for more than 10 years need less IT assistance to access data from the E.P.R. system than other user groups do. To test this hypothesis, this thesis will use question 25 as the independent variable and question 11 as the dependent variable. Question number 25 groups users by asking them about their age range, while question 11 asks the subjects about their level of dependency on IT assistance to access data from the E.P.R. that they cannot get on their own. Based on the collected responses, there was a statistically significant difference between groups as determined by one-way ANOVA ($F(3,21) =1.7954$, $p = 0$).

A Tukey post-hoc test revealed that the dependency on IT assistance was statistically significantly lower with senior groups of 11 to 15 years of experience (2.00 ± 0.707, $p = 0.001$) compared to the less experienced with 3 to 5 years (4.33 ± 0.577) and 6 to 10 years (4.67 ± 0.816). None of the interviewed subjects had less than 3 years of experience.

The third hypothesis is that doctors access the E.P.R. system from outside their work premises more often than other users. The test for this hypothesis needs question 26 as the independent variable and question 14 as the dependent variable. This thesis grouped the subjects based on their highest achieved degree, and this thesis asked them if their work requires that they access or connect to the hospital’s data from outside the hospital.

There is an internal procedure for accessing the E.P.R. data through a VPN connection to the hospital's servers. This access is provided by the IT department based on the written consent of the management.

The question is not for testing the ability of the user to connect to the workplace, but rather to test the need for this user and whether they actually use this access or not. Based on the collected responses, there was a statistically significant difference between groups as determined by one-way ANOVA ($F(3,21) = 1.769$, $p = 0.037$). A Tukey post-hoc test revealed that accessing the E.P.R. system from outside the work premises was statistically significantly higher with doctors' group (2.25 ± 0.957, $p = 0.046$) compared to the bachelor's degree owners (1.00 ± 0.000). There was no significant relationship between the other groups.

5.3.3 Qualitative Analysis

To properly describe the problem and analyse the results of the case study, a comprehensive methodology needs to be applied and maintained throughout the analysis process. The Soft-Systems methodology (SSM) is the most suitable to use in similar problem analysis processes, because it allows to descriptively and visually identify the actions to be take and is most appropriate with problems where the desired solution cannot be expressed in concrete technical formats. SSM defines a set of tools that can be used to initiate communication with stakeholders and clinicians and express the needs and describe the current situation in the form of diagrams that can be later implemented by software technologies.

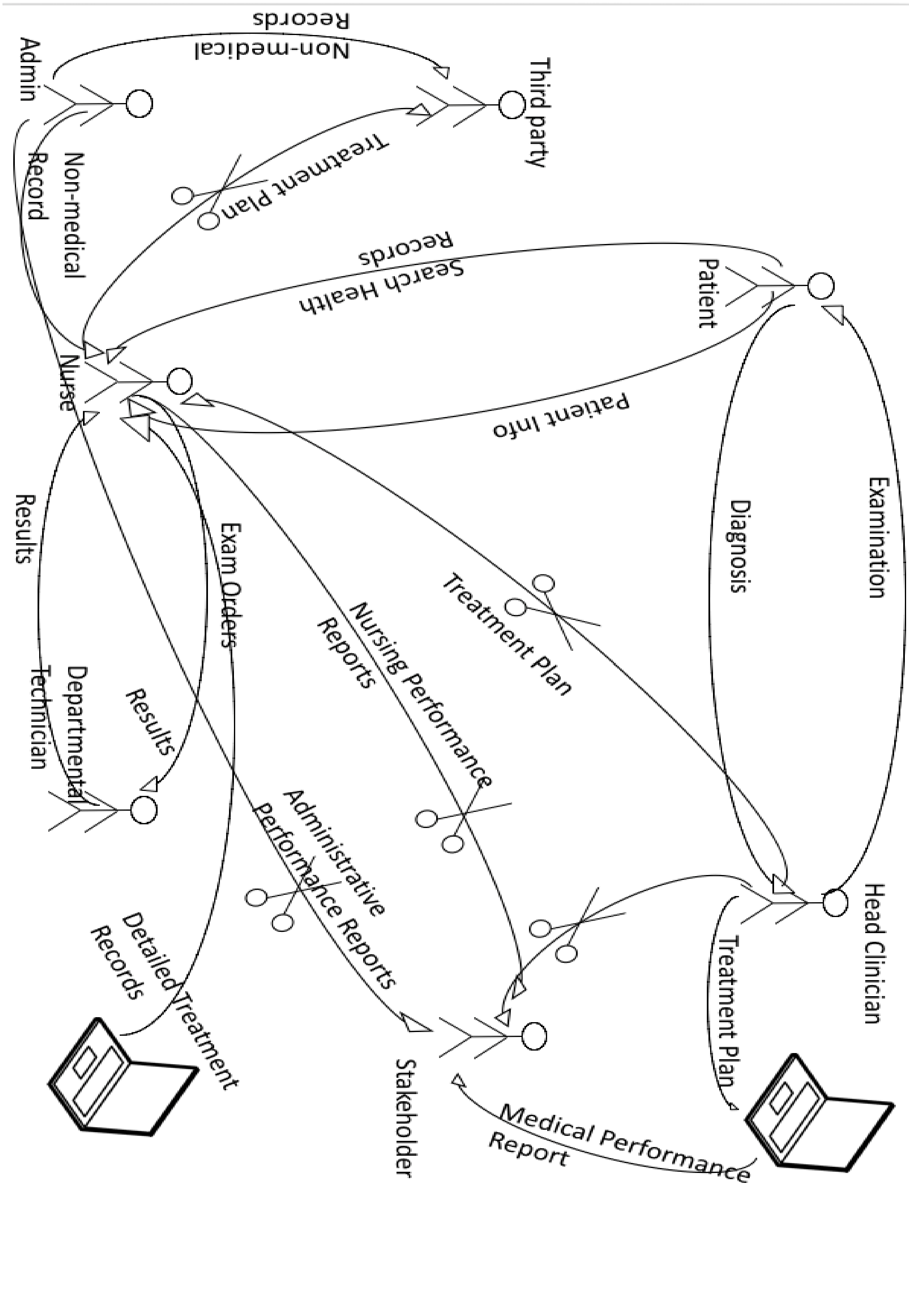

Figure 5.6: The RICH picture of the current situation

The RICH picture shows conflicts in interest between three major parts of the organization; the doctor/stakeholder relationship, the decision-support/stakeholder relationship and the decision-support/staff relationship. This thesis will develop the conceptual models of the EPR environment by stating the root definitions and the CATWOE diagrams that explain how the system and its constituting sub-systems operate.

Root Definition	Relevant to healthcare organization EPR system
The organization maintains patient information in the Electronic Patient Record by collecting and analysing data from patients and health specialists in order to achieve better quality of care and improve the performance of the organization.	

Table 5.1 Root Definition of the healthcare organization EPR system

CATWOE

Client	Patient, who receives the services of the healthcare providers
Actor	Doctors and nursing mangers who interact with the patient during the encounters within the healthcare setting
Transformation	Need to deliver healthcare services to patients efficiently and timely; that need is met
Weltansaung	Information about the patient needs to be stored in the EPR systems and used in the process of delivering the healthcare services to the patients.
Owner	Stakeholders and Healthcare providers
Environment	Healthcare setting

Table 5.2 CATWOE of the healthcare organization EPR system

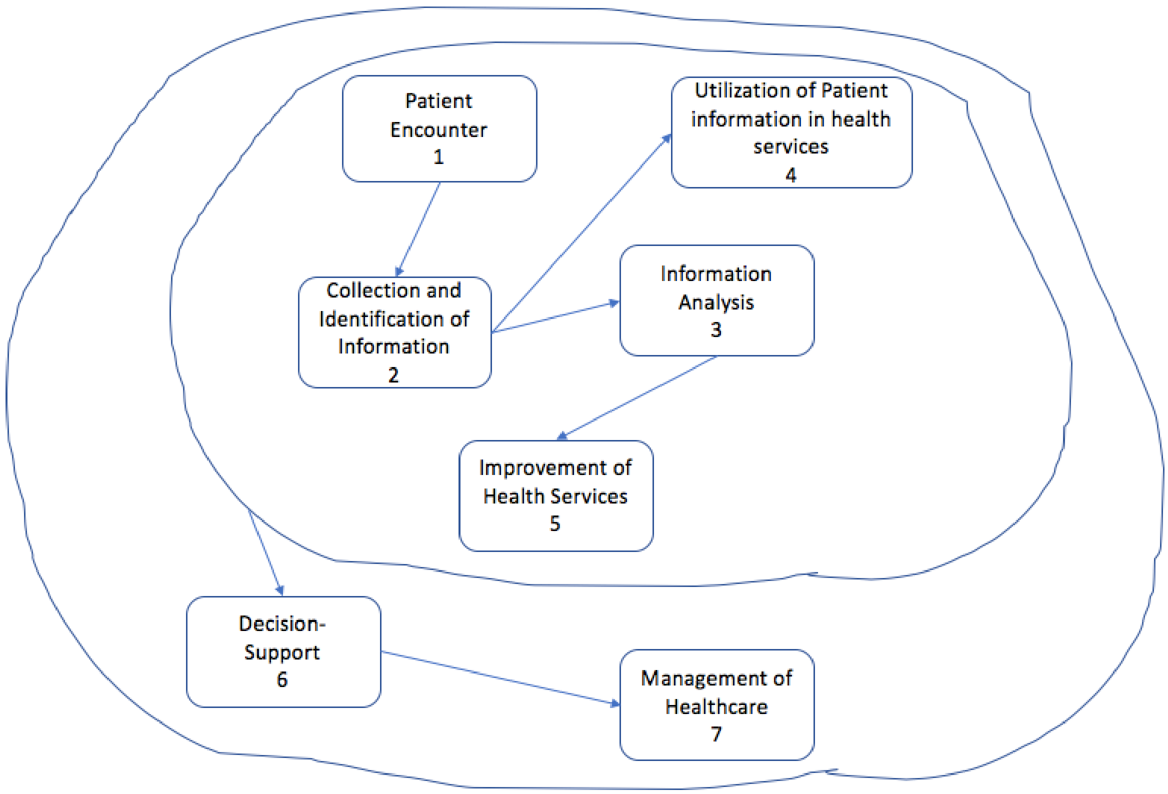

Figure 5.7: Conceptual Diagram of EPR system

Root Definition	Relevant to the improvement of processes by Stakeholders
The Stakeholders in the healthcare environment improve the healthcare process by making decisions that change the organization in order to improve the quality of care for patients.	

Table 5.3 Root Definition of the stakeholders and decision-support subsystem

CATWOE

Client	The healthcare organization, that needs to be changed by decisions
Actor	Stakeholders, who make decisions based on the needs for change
Transformation	The healthcare organization is changed by the decisions of the Stakeholders; that need is met
Weltansaung	The healthcare services provided by the organization are continuously improved by the decision made by the Stakeholders and result in better quality of care.
Owner	Patient, who receives the healthcare services
Environment	Healthcare setting

Table 5.4 CATWOE of stakeholders and decision-support subsystem

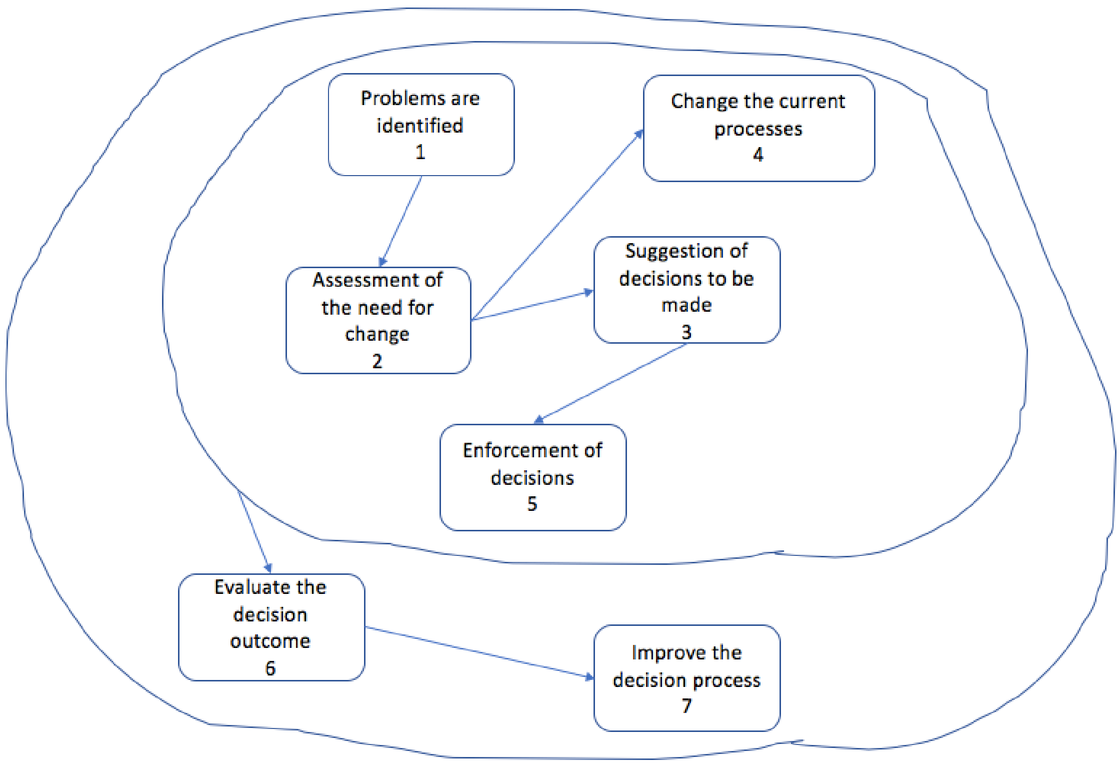

Figure 5.8: Conceptual Diagram of the stakeholders and decision-support subsystem

Root Definition	Relevant to the treatment of patients by doctors
Doctors treat patients at healthcare centres and clinics by examination and prescription of medication to heal and improve patient health	

Table 5.5 Root Definition of the treatment of patients by doctors' subsystem

CATWOE	
Client	Patient, who receives the treatment from the doctors
Actor	Doctors who supervise the treatment of patients through prescription of medications and follow-up on their health improvement.
Transformation	The patient needs to be treated and healed that need is met
Weltansaung	Patient information is collected by the doctor at the encounter and stored in the electronic patient records of the patient to contribute to a better treatment in the future for other patients.
Owner	Healthcare centres and clinics

Environment	Healthcare setting

Table 5.6 CATWOE of treatment of patients by doctors' subsystem

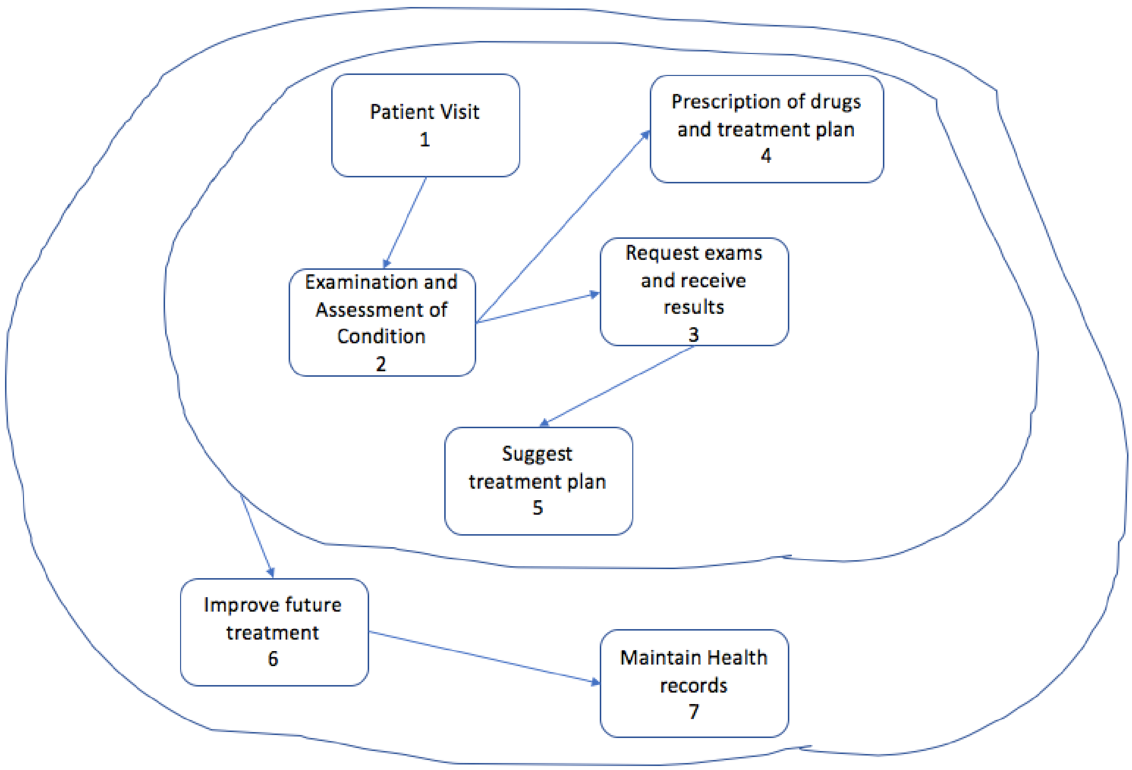

Figure 5.9: Conceptual Diagram of treatment of patients by doctors' subsystem

Root Definition	Relevant to the collection of patient data by staff
Nursing staff and managers record information about patients by collecting this information from patients to expand the electronic patient records and achieve better quality of care	

Table 5.7 Root Definition of the collection of patient data by staff subsystem

CATWOE

Client	Patient, who's data is collected and stored
Actor	Nursing staff and managers who get the information from the patient, and decision-support officers who process and analyse this information
Transformation	The healthcare organization needs the data to be collected by the nursing staff and managers □ that need is met
Weltansaung	The information that is collected about patients is stored in the electronic

	patient record and is used to improve the quality of care through analysis by decision-support officers and managers
Owner	Healthcare centres and clinics
Environment	Nursing stations and points of care

Table 5.8 CATWOE of collection of patient data by staff subsystem

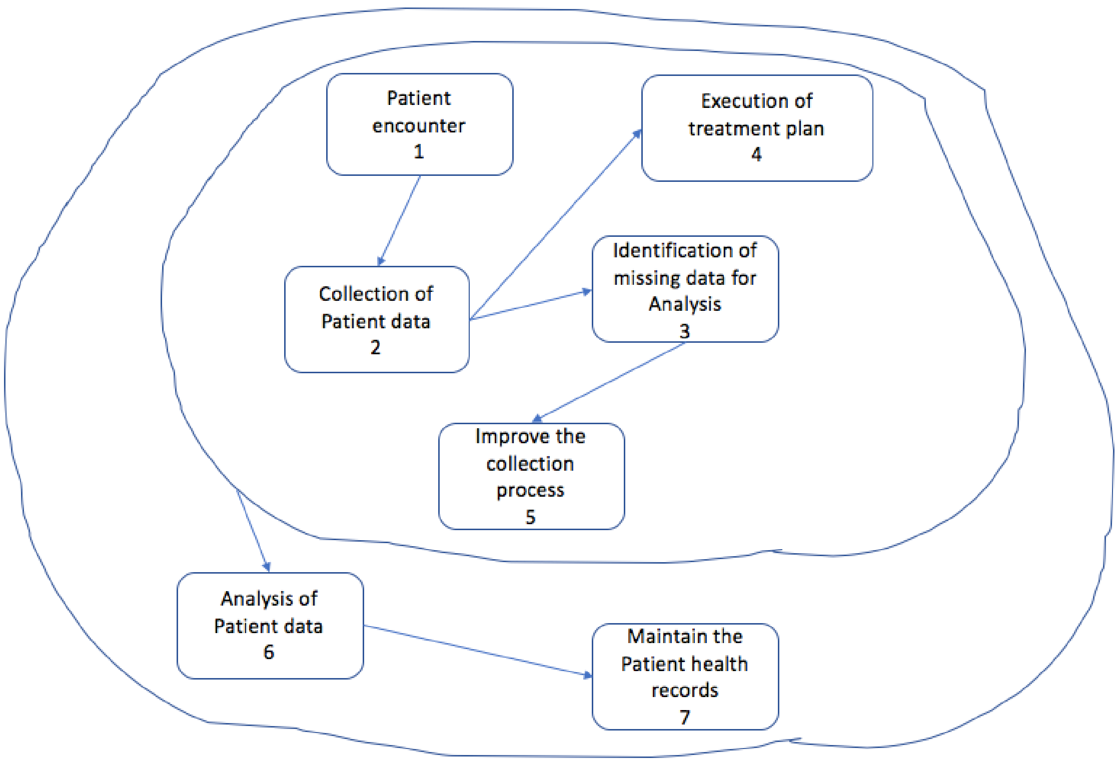

Figure 5.10: Conceptual Diagram of collection of patient data by staff subsystem

5.4 Requirements Analysis of the Framework System

5.4.1 Functional Analysis

The main activities are shown in the Root Definition of the patient treatment and encounter and in the conceptual model. The setting is the classical healthcare location, where the patient undergoes several diagnosis and health assessment activities before a treatment plan is put to execution and the assessment is later performed on the patient to track the health progress, leading to either a successful treatment or a failed one. During this encounter, the documents that are related to the treatment and health records of the patient are populated and analysed by different functional areas and actors in the system. The following hierarchical structure of the functional areas and their relationships is the starting point for the section that follows, where the structure of the system is analysed and further explored.

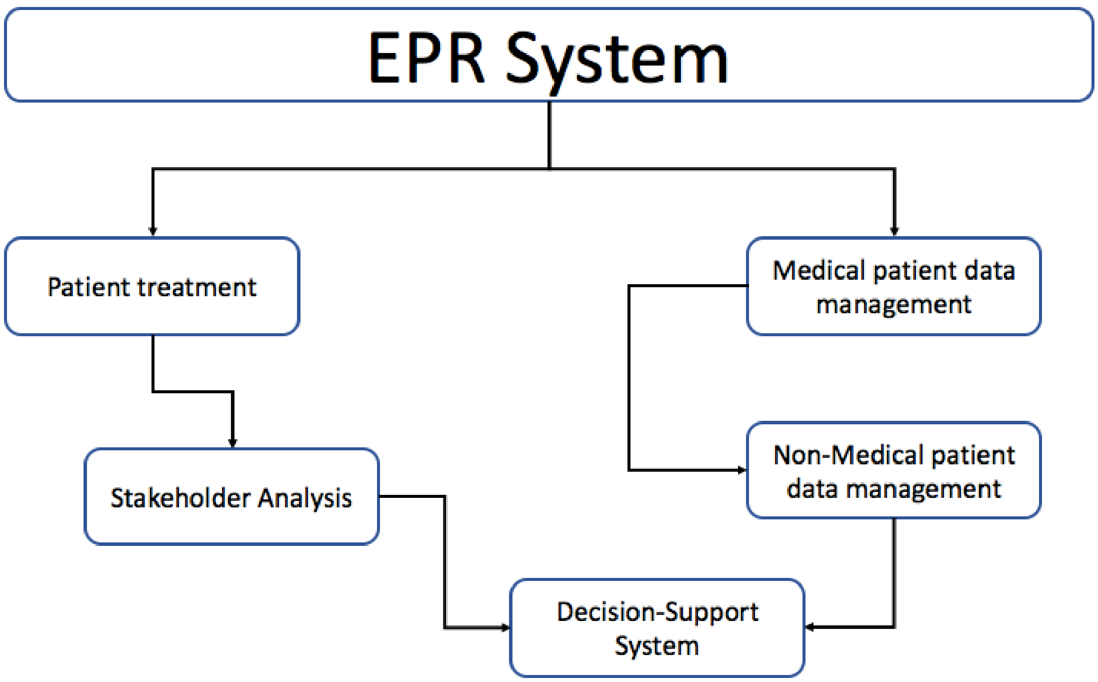

Figure 5.11: The Functional Decomposition Diagram

5.4.2 Non-functional Requirements

In addition to the functional requirements, the discussions and meetings resulted in a list of non-functional requirements (NFR's) and expectations from the framework system to be developed. These requirements and expectations were categorized and listed in the table below:

Category	Non-Functional Requirement
Functionality	Ability to access multiple data sources

Usability	Work on multiple forms at the same time
Usability	Reduce transitions from one form to another by using composite layouts
Efficiency	Eliminate repetitive actions through template-based form design
Usability	Ability to save the current process and continue it later
Usability	Integration with existing Single-Sign-On (SSO) providers to reduce logins
Usability	Allow multiple themes to customize the look-and-feel
Functionality	Having a customizable user dashboard
Usability	Connect to online data sources (security constraints apply)

The collected NFR's reflect the users' needs to produce quality output to the monthly status report under construction and would therefore need to ensure that the framework system allows them to work more efficiently and reduces the manual tasks that they perform. In addition, the users want to ensure that the framework system is easy to use and does not require any technical knowledge. They have previously used some commercial productivity tools and would need to find similar functionalities inside this tool, at least for the purposes of authoring new data models or initiating data uploads.

5.4.3 Knowledge Analysis

At this point the entities and rules of the current situation study are presented, which would then contribute to the class-diagrams in the data analysis section.

The entities in the system are the following: head clinician, nurse, stakeholder, patient, third-party agent, system admin, departmental technician, patient health records, decision-support reports, non-medical patient records, administrative performance reports, the treatment plan.

The list of rules governing the entities is the following:

- The head clinician assesses the patient's medical condition and orders a treatment plan.
- The nurse follows up on the patient's health condition and continuously updates the health records. The nurse also ensures that the treatment plan is executed and also performs the needed communication with other identities in the system to reach a conclusion of the treatment.
- The stakeholder is the organization's top management and needs concise and accurate information that has been properly filtered to present a decision-support.
- The patient is the owner and producer of the data in relation to his/her health.
- The third part agent interacts with the healthcare setting from an external location/process and needs information from the healthcare environment to make decisions related to the patient's financial status, safety measures, quality of service and other statistical enquiries and studies.
- The system admin ensures the proper and continuous operation of the healthcare information system.
- The department technician receives patient-related samples or records and analyzes them using manual or automated examination techniques to produce a set or results and diagnoses.
- The patient health records contain all information and data related to the health status and medical treatment of the patient at different points of care.

- The decision-support reports enable stakeholders and other top management decision makers to base their decisions of improvement and patient treatment techniques and strategies on reliable and solid data coming from various analytical executions against the existing data pools.
- The non-medical patient records are data related to the patient which is not medical. This could include the financial, social and other relevant information.
- The administrative performance reports help the management evaluate the overall status of a well-defined section of the healthcare processes in a continuous and validated manner. Examples could be the occupation rate reports, the recovery rate reports, the time-to-heal average reports, and others.
- The treatment plan is a set of steps, measurements, procedures and other executable activities that need to be performed in sequence by healthcare professionals to ensure the proper treatment and healing of a patient. This could include both medical and non-medical activities. The treatment plan is usually contained within the hospitalization process but could sometimes extend beyond it.

5.5 Data Model of the Framework System

The conceptual design that this thesis proposed requires that a modified data analytical framework be implemented that solves the limitations presented in Section 4.3. To do this a data model that encapsulates the required process and makes it possible for pre-loading data quality assurance, parameterized analytical algorithm execution with the framework, and easier data consumer interaction with the framework has been created. This section will show the data model diagrams below and explain the core data model classes in detail, how it is associated with the other model classes, and the purpose of this association from the process perspective.

The first class-diagram is the “Batch” diagram. This diagram shows the relationship between the “Batch” class and the mapping classes – DataMap and DataSet. The User class is also shown because a Batch always has an owner, who is the user who created the batch entry. The purpose of loading data in batches is to allow for marking the batch as “enabled” when it is ready to be used in analysis. Including and excluding data batches in analytical results also allows for “test-runs”, which give the clinicians the ability to test their algorithms for new data. The DataMap class is the header construct for the data mapping mechanism that is implemented, during which the system converts incoming data to performance entries.

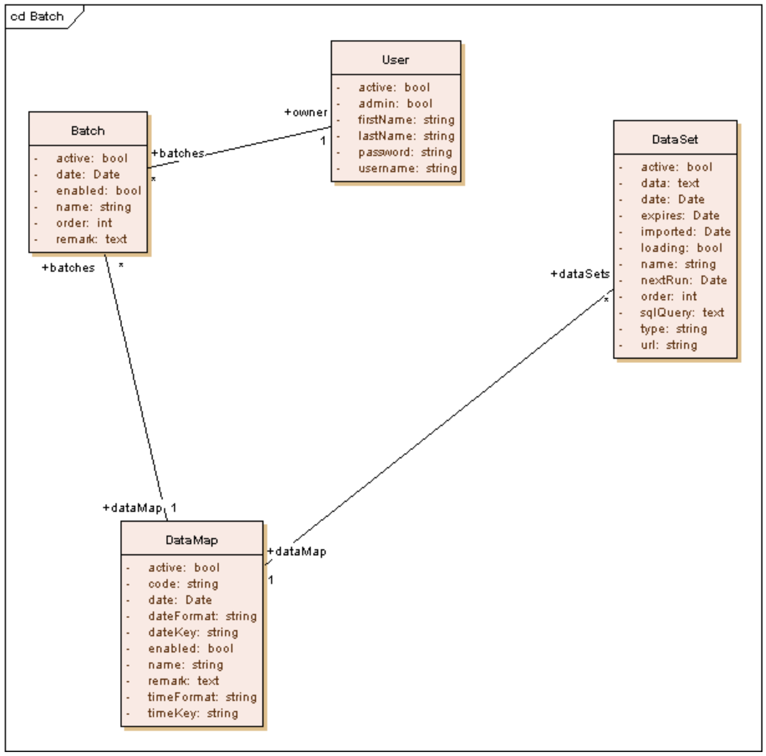

Figure 5.12: The Batch class-diagram

The second class-diagram describes the data model for the data-loading process. The data is loaded from multiple data-sources, and these should to be properly defined as records in the system to enable scheduled loading. The scheduling mechanism is controlled using intervals, and This thesis use a mechanism like the scheduled-job design pattern where the next fetch date is stored within the Data-Set. The mapping mechanism is shown here as part of the loading process, where the incoming data is consumed as a tabular structure, and the DataSet retrieving this data delivers the content to the mapping activity for processing. The mapping is applied by specifying the DataField entries that the data represents, where those data fields could represent an entity in the system such as doctors, nurses, drugs, operations, or any other data producer or activity participant. A data field also represents the values of the performances, which could be alphanumeric. The performance is captured at a specific date/time and it is therefore important that the tabular columns containing the date/time content be defined in the data map fields dateKey and timeKey.

The identities in the system are referenced by code-name pairs and are tracked by the Identity class which represents the performance data-producers. For this reason, the DataField class is associated with the Group class, because every Identity in the system belongs to a Group. Finally, and to complete the performance capture mapping activity, the DataMapValue class is introduced to load the performance values based on the Measure applied. For example, it could be the case that the performances of three measures are being captured, so in this case there will be three DataMapValue entries for the DataMap of the data capture.

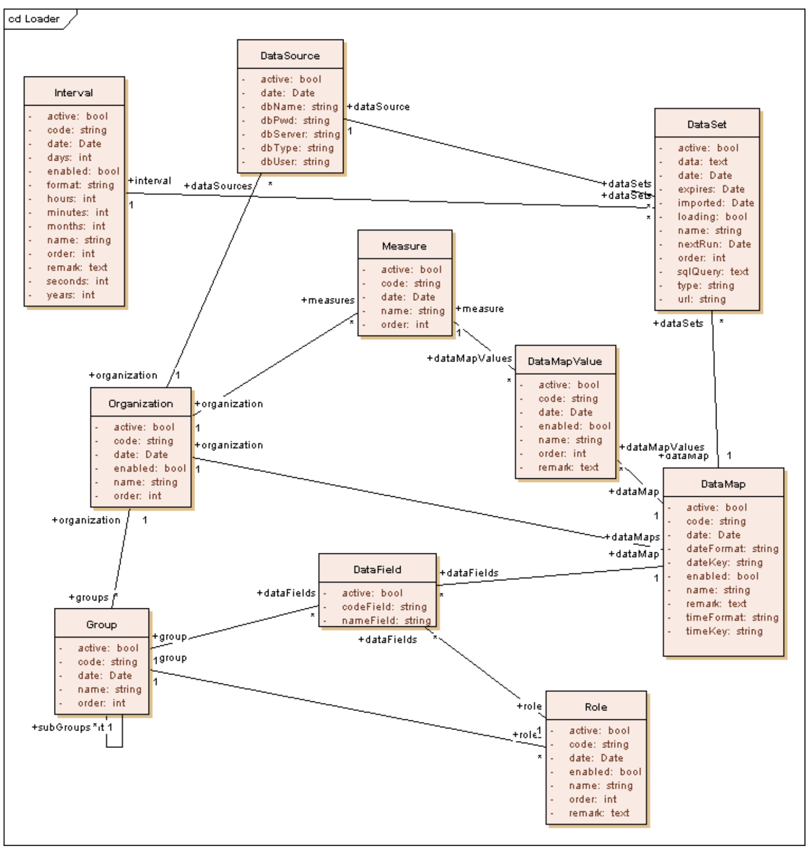

Figure 5.13: The Loader class-diagram

Another structure that is key to the data-model is the Identities class-diagram, which defines the relationships and associations related to the Identity/Group design-pattern. In many loaded performances, the different identities from the same group might assume different roles, as is the case in an admission transaction where the doctor group is found in two participations; the admitting and the admission referral doctor.

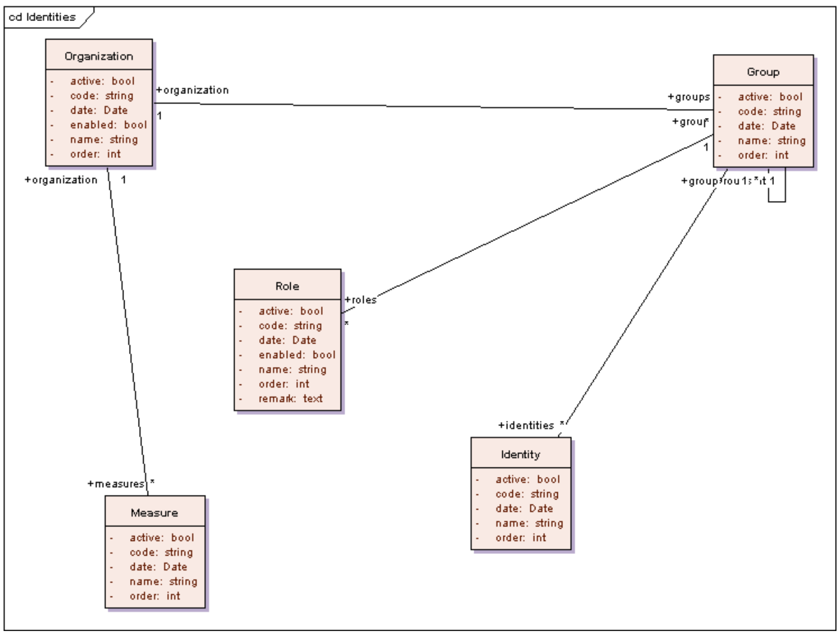

Figure 5.14: The Identities class diagram

The interaction of clients with the framework application is handled through a user design-pattern, where the access is not only restricted to user authentication but is also based on visibility criteria for the Organizations, Groups and Measures. This allows system administrators to segregate users based on the required data visibility and ensures that algorithms created by users are filtered properly to exclude non-visible data. The three visibility criteria govern the availability of data for analytical processing throughout the system, favouring activity-based security to the classical user-based security. The framework allows for multiple organizational data hosting, and therefore the construct of Organization can serve as a top-level data divider.

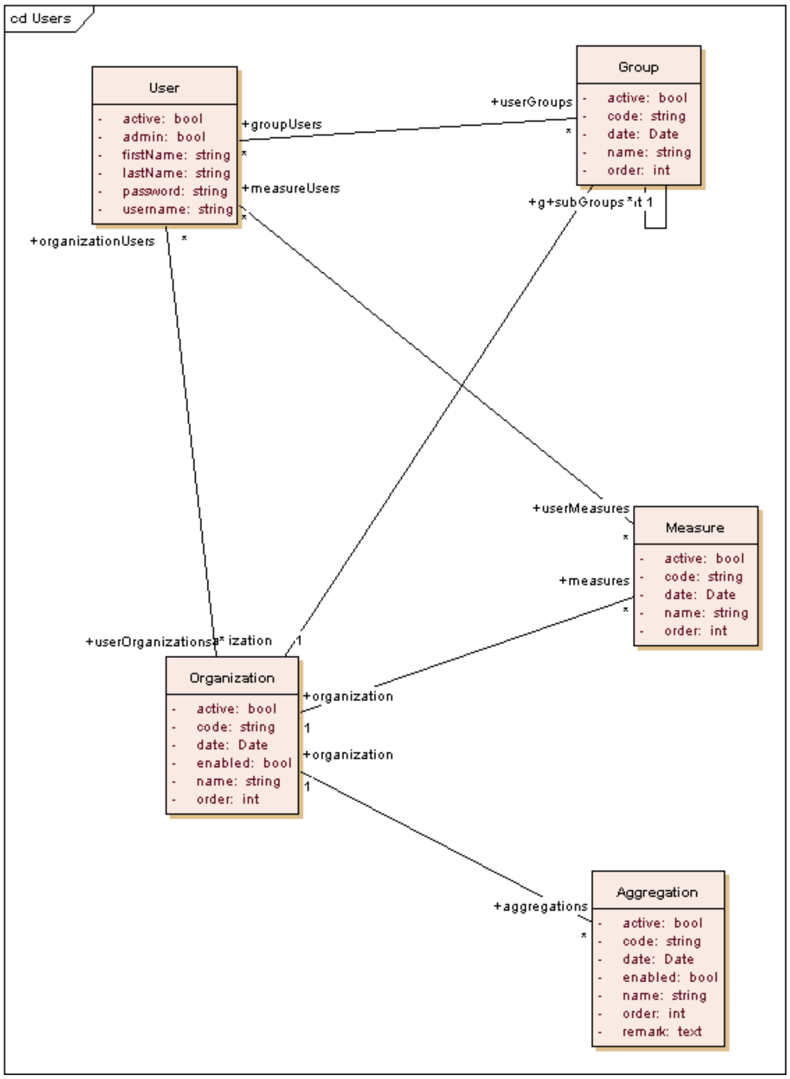

Figure 5.15: The Users class-diagram

The application represents loaded data from the different data sources as performance records. The performance measurement and management (PMM) concept is implemented through this data-model in the application, and the abstraction of identities enables the compression of the data-model to a single performance-oriented approach. The performance records are created following the application of the transformations defined by the data map. The Performance class encapsulates the top-level performance header record and is related to a Batch, as explained in the discussion of the batch class-diagram. The PerformanceIdentity class links performances with identities in a many-to-many association, but also specializes the identity association with a role definition when required. In the example presented above, the admitting and consulting doctor roles clarify the presence of two identities in a performance record. The DataSet identifies the performance record as part of a bigger data-set, and the batch marks the performance for association with a loading activity. The PerformanceValue class is where the multiple values from different measurement records are saved and is therefore associated with the DataMapValue class to identify the mapping that produced it.

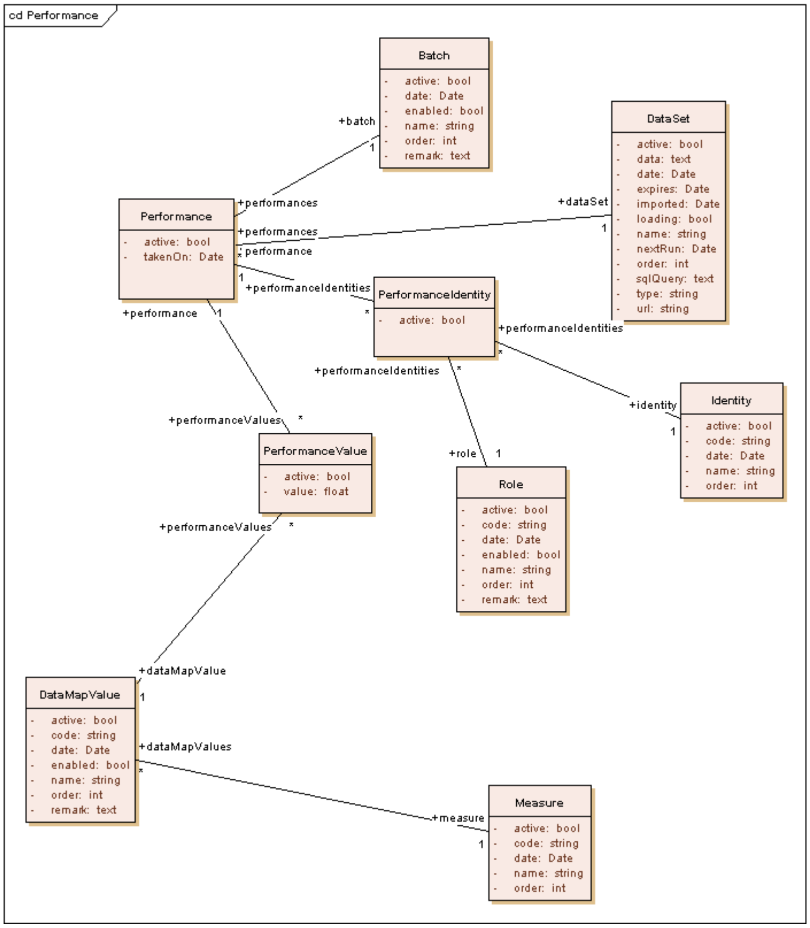

Figure 5.16: The Performance class-diagram

The loading process is enhanced with a transformation mechanism that allows the data-owners to manage and maintain the loading of their data. This comes in the form of a transformation definition data-model, where the mapping activity is altered at load time by a series of change steps that are applied in order on the performance values. The Operation class includes a set of pre-defined operations ranging from basic manipulations to advanced scriptable functionalities. This makes it possible to include custom parametric transformation operations that can be easily re-used and to cascade mapping transformations. The “parameters” attribute in the ChangeStep class serves as an input to the “script” attribute in the Operation class.

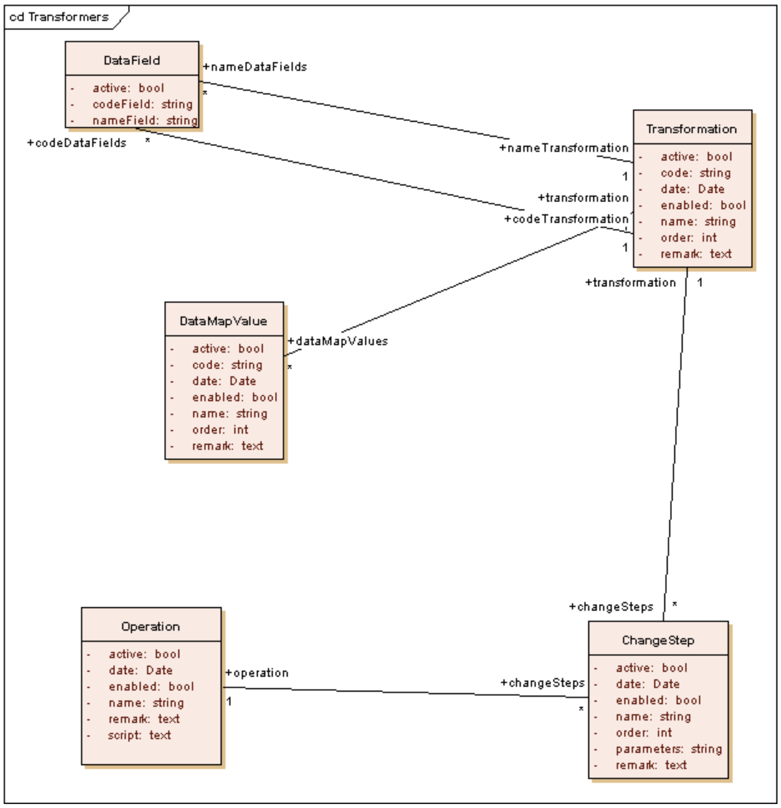

Figure 5.17: The Transformers class-diagram

5.6 Summary, conclusion and discussion

This chapter performed a quantitative and a qualitative analysis of a case study that was conducted in a local healthcare setting. The purpose of the case study was to gain more visibility into healthcare setting and then produce a picture of this setting that is validated against the case study results. The information gathered was then statistically analysed to conclude the validity of hypothetical statements from previous chapters. The chapter analyses the requirements of the roles from different healthcare identities, and then creates a data model to represent a knowledge information system that helps fill the gaps and limitations encountered in this and previous chapters.

Chapter 6

Design and Development of the Framework System

6.1 Introduction

In this chapter, the proposed implementation stages and steps for building the clinical analytical framework are presented. The chapter is divided into two main sections. The first section discusses the methodology that This thesis adopted to design the framework solution. The software methodology is presented in moderate detail to allow reproducing the framework by peer researchers should that be needed. The methodology allows to build the framework by first designing it and then moving the models to the production stage. In Section 6.6 the framework implementation is explained, first by transforming the prepared design in Chapter 5 and implementing the methodology of UML from Chapter 2, and then by creating the adaptations that are needed to make the system ready for testing. The framework is built using modern web technologies and adapts some software engineering design patterns to improve its flexibility when used in clinical analytical environments.

6.2 System Architecture and Modular design

The remaining sections in this chapter will present a deep explanation of how the framework system is designed and developed to meet the requirements from Chapter 5, but before this task is undertaken, a top-level view of the system's architecture and how its modules interact with each other to fulfil these requirements is explained in this section. The discussion in this chapter follows the actual data from its initial source systems within the healthcare setting until it reaches the targeted users through a web-interface and a portal web-application.

The framework system's architecture follows the service-oriented architecture design, where the entire system is broken down into consuming and producing modules that co-exist to deliver services and data to each other through a multi-layer model, as shown in the figure below:

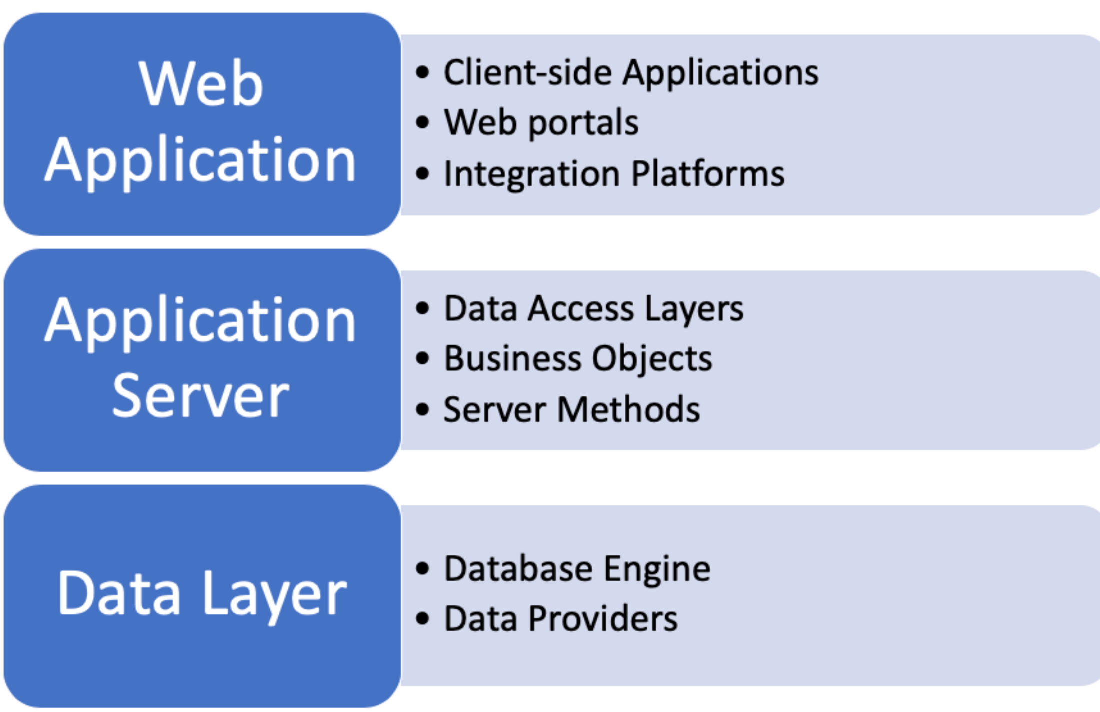

Figure 6.1: Service-oriented architectural design

The framework system in this thesis has many modules that contribute to the solution of the problem identified earlier in previous chapters. The general structure of the proposed solution process in Figure 4.2 exposes several modules that would be required to operate and execute as part of the architecture design from Figure 6.1, and therefore these modules' locations and functional dependency is shown in the following figure, which serves to explain the topology and helps guide the discussions in the next sections of this chapter:

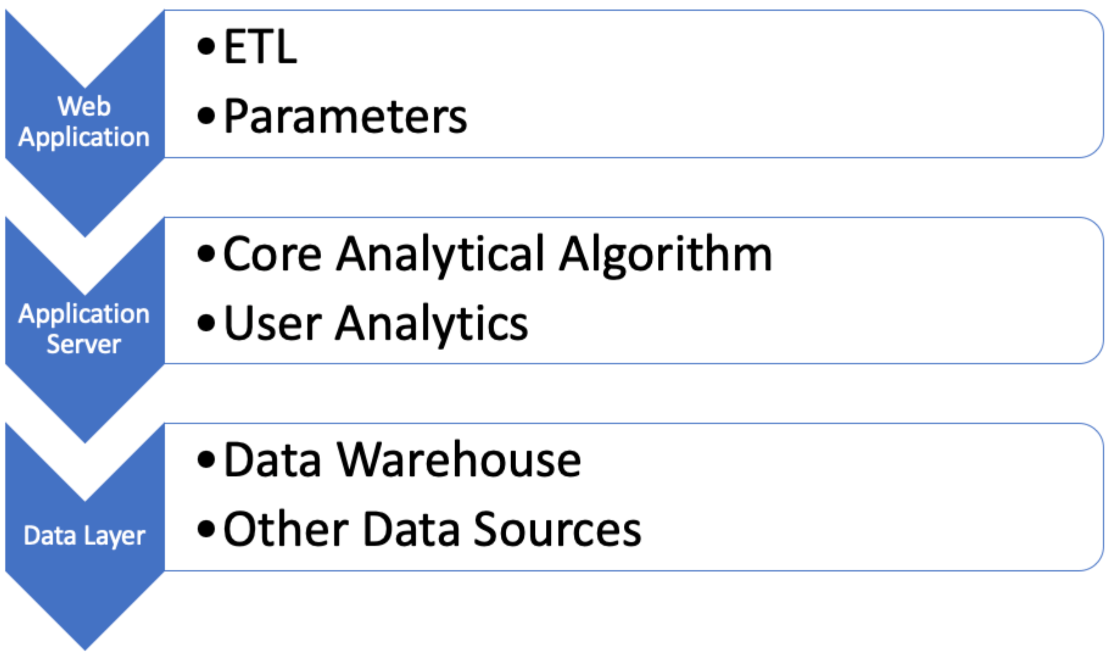

Figure 6.2: Framework Modules as part of the Architectural Design

As the data is extracted from external healthcare systems and devices, the user applies transformations to this data within the web-application layer and forwards these new transformed and restructured data sets into the application server layer. The data extraction may also be parametrized, for example using date ranges or segments. The users may also define their analytics inside the web application, but those analytics are part of the application server layer, unlike the ETL implementation which is completely client-based. Eventually the application server executes the user analytics as part of the core analytical algorithm which will be discussed in detail in the following sections of this chapter. At the basis of this execution is the utilization of the loaded data and data models from the data layer, served through the data abstraction layers and business objects shown in Figure 6.1.

6.3 Feasibility of the Framework System

The proposed implementation of an analytical framework where dynamic meta-data schemas are applied to loaded data from different data sources faces multiple challenges. The solution proposed in this thesis would resolve the gap found in the literature review and in the case study analysis but would require an effort related to building a complex algorithm that would apply a dynamic enquiry built by a framework user to a dynamically loaded data from one or a multiple set of data sources. In addition, the imported data could also come from de-normalized data aggregations. The system should allow the user to control the data loading procedures while ensuring data integrity, in addition to enforcing data safety and privacy principles through an identity-based visibility and control privileges model. The main focus of the proposed framework is to allow non-technical clinicians and data analytics professionals to reduce the time spent to generate their desired decision-support outcomes through a self-serving method of moving the data analytics customization and planning to the side of the data owner and output receiver. The challenge in this model is the ability to work through a dynamic meta-data model that is based on minimal or non-existing assumptions regarding the source data and the transformation mechanism and leading to a dynamic output that is not constrained by the functionalities of the framework design.

6.4 The Dynamic Analytical Algorithm

At the core of the proposed framework is an algorithm that is designed to handle the dynamic nature of the user interactions and customized data models embedded within this framework. The details of database-access would be presented in later sections of this thesis, but in this section the abstract algorithm for retrieving data and reconstructing knowledge from dynamic user queries is presented.

The Dynamic Analytical Algorithm (DAA), hereafter referred to as the algorithm, starts with retrieving the structure of the desired analytical query that the user developed. This is achieved by processing the steps of the enquiry from the data model, and with every step the algorithm converts the query step to an appended filtration against the performance data model which encapsulates the loaded data into the framework's data warehouse.

Performance data structures are represented in the data-model with three core classes. The Performance class, which holds the header definition of every row-level entry in the loaded tabular structure. The Performance class is referencing the Batch class, which allows the loader module to mark performance records and source them back to a particular batch load. The performance class is referenced from both the PerformanceIdentity class and the PerformanceValue class. These two classes hold the details of the performance record and would therefore be queries based on the user enquiry in question.

The algorithm can be used in two ways. The first one is that the enquiry parameters, the steps of the use enquiry, could be already stored in the database. This is particularly useful when a user wants to re-run a saved enquiry or append to an existing one. The second way is through an ad-hoc enquiry, representing an analytical query that is not stored in the database, and is therefore the result of the user's interaction with the system objects. This is sometimes useful when users want to inspect the existing knowledge warehouse before they start running their actual queries. In both ways, the algorithm ensures that the underlying steps are the same. To do that, the algorithm looks for any ad-hoc parameters that might be provided, and if none are found, checks the stored parameters in the database, as shown below:

```
EnquiryParameter[] eps = ee.EnquiryParameters;
if (eps == null || eps.Length == 0) eps = new EnquiryParameter {
 Active = true,
  Enabled = true,
  _EnquiryExecution = ee
}.findAll();

```

Figure 6.3: Identifying the query steps source

The algorithm starts by checking the EnquiryParameter records of the provided enquiry. If there are none, it will look for the already stored ones in the database based on the query that the EnquiryExecution object represents. The EnquiryExecution class serves as a header to the enquiry parameters. The findAll() function from the data-access class model that is produced from the UML class diagrams is used, and it will convert the object query (represented by the eps variable) to an array of EnquiryParameter objects. At this point, all information related to the user's desired query is available. The algorithm then proceeds with creating the generic Identity object that will represent the output of the algorithm execution. This is done by ensuring that the user enquiry spans only one group of identities. This is very important, because the data segregation requirements are such that information is always constrained to a particular classification of objects. If the user enquiry does not explicitly indicate a grouping of the desired identities, the algorithm restricts the output to only those identities that match a dynamic grouping filtration, which is presented in the figure below:

```
Identity id = new Identity {
 Active = true
};
try {
 // a stored enquiry would have a group
 id._Group = ee._Enquiry._Group;
} catch {}

if (id._Group == null) {
 // an ad-hoc query: deduce the group indirectly
 id._Group = new Group {
  Active = true,
   Enquiries = new Enquiry[] {
    new Enquiry {
     Active = true,
      Enabled = true,
      EnquiryExecutions = new EnquiryExecution[] {
       ee
      }
    }
   }
 };
}
```

Figure 6.4: Restriction of output Identity results

The algorithm forces the output Identity results to be all associated with a group that the EnquiryExecution query covers.

The next step in the algorithm execution is based on going through all of the EnquiryParameter steps and construct the output Identity filtration based on what those steps instruct. This is achieved first by looping over these parameters and preparing the Performance object that will match the query step:

```
Performance last = null;
foreach(EnquiryParameter ep in eps) {
 Performance p = new Performance {
  Active = true,
   _Batch = new Batch {
    Active = true,
     Enabled = true
   }
 };

 if (ep.StartDate > User.getMinDate()) {
  p.TakenOn = ep.StartDate;
  p.setFindComparison("takenOn", "BETWEEN", ep.EndDate);
 }
 p._DataSet = new DataSet {
  Active = true
 };

```

Figure 6.5: Preparing the Performance object to match the step filtration

The algorithm allows that a time-framed part of the overall Identity matches be filtered in each step, and not in a global context. It could be the case that the user required to look into all Patients that have had an operation during the last year. The time restriction applies to the operation, not to the patient admission. In that sense, the patient might have been admitted to the hospital 2 weeks before the start of last year but had to undergo certain medical treatments leading to the time of his/her operation, sometime during last year. This allows the user a good level of time framing at every step of the enquiry.

The algorithm restricts the desired performance records to only those that the enquiry step associates with a measurable metric. Going back to the example of the operations of last year, the user might need to look at operation durations, therefore filtering only by operations that consumed more than 3 hours.

```
if (ep._Identity != null) {
 p.PerformanceIdentities = new PerformanceIdentity[] {
  new PerformanceIdentity {
   _Identity = ep._Identity
  }
 };
 if (ep._EnquiryCondition != null && ep._EnquiryCondition.IsNot) {
  p.setFindComparison("performanceIdentities", "NOT IN");
 }
} else if (ep._EnquiryCondition != null) {
```

Figure 6.6: The case when the EnquiryParameter is already filtered to a particular Identity match

At this stage of the step analysis there are two possibilities for the desired related identities in question. The EnquiryParameter step might be focused to a particular filtration of the desired identities. In the example presented above, the user is interested in patients that have had surgery during last year. The surgery is an operation Identity in the loaded data of a healthcare organization and would thus be defined in a certain catalog or tabular structure. The user might want to focus the filtration of the operations to a particular list, or even to a set of operations that span a particular subset of the entire operations list. It is worth noting at this point that the Identity association from the EnquiryParameter step would therefore allow the linking of multiple user enquiries. In our example, another enquiry might have already been created, possibly by another user, that delivers the list of operations that require special anesthetic procedures. As such, the user's patient enquiry could re-use this base enquiry in this step. The algorithm also supports negated filtrations, as shown in the checking for the "IsNot" attribute of the EnquiryCondition association of the EnquiryParameter step.

It is also possible that the user has not applied any filtration to the related identity of the enquiry step under processing. This is handled in the logic presented in the figure below:

```
59  } else if (ep._EnquiryCondition != null) {
60    p.PerformanceValues = new PerformanceValue[] {
61      new PerformanceValue {
62        Active = true,
63        _DataMapValue = ep._EnquiryCondition._DataMapValue
64      }
65    };
66    string having = ep._EnquiryCondition._Aggregation + "(value) ";
67
68    if (ep._EnquiryCondition.IsBetween) {
69      having += "BETWEEN " + ep.Value + " AND " + ep.MaxValue;
70    } else {
71      string comp = "";
72      if (ep._EnquiryCondition.IsGreater) comp += ">";
73      if (ep._EnquiryCondition.IsLess) comp += "<";
74      if (ep._EnquiryCondition.IsNot) comp += "!";
75      if (ep._EnquiryCondition.IsEqual) comp += "=";
76
77      having += comp + ep.Value;
78    }
79    p.having(having);
80    p.groupBy("identitieid");
81  }
```

Figure 6.7: The case when the step is not filtered for the underlying identity

Here the algorithm will attempt to determine the underlying identity to be filtered against, based on an aggregation of performances. The aggregation method is prepared to be applied to the “value” of the performances, which could either be spanning a date range, as in the case of the condition in line 53, or as a normal value comparison based on classical comparison operators. The purpose of this part of the algorithm is to allow users to filter the list of resulting identities based on their performance in a particular measure. In the example of the operations from last year, the user might want to filter the patients to a subset that underwent operations that lasted more than the average of normal anesthetic operations from a 3-year range.

As the algorithm’s main loop goes over the parameters of the enquiry, the algorithm keeps track of the last Performance that was generated from the parameter in the current loop.

```
if (last == null) {
 last = p;
 continue;
}
```

Figure 6.8: The algorithm keeps track of the last Performance object

Within every loop, the algorithm utilizes a filter object of type Performance that establishes a filtration mechanism linking the performances of the enquiry parameter within the current loop with the main Identity filter.

```
Performance filter = new Performance {
 PerformanceIdentities = new PerformanceIdentity[] {
  new PerformanceIdentity {
   _Performance = p
  }
 }
};
```

Figure 6.9: The filter object within the loop of the algorithm

Within the main loop of the algorithm, the filter is further populated and restricted to a subset of the Identities of the group defined by the parameter.

```
if (ep._EnquiryCondition.WithOccurrence) {
 object[] data = null;
 p.Id = -1;

 if (ep.Occurrence > 0) {
  // preceding the previous takenOn date
  data = new object[] {
   filter,
   TimeSpan.FromMinutes(0),
   filter,
   TimeSpan.FromMinutes(ep.Occurrence)
  };
 } else {
  // following the previous takenOn date
  data = new object[] {
   filter,
   TimeSpan.FromMinutes(ep.Occurrence),
   filter,
   TimeSpan.FromMinutes(0)
  };
 }

 last.setFindComparison("takenOn", "BETWEEN", data);
} else {
```

Figure 6.10: Filtration based on the occurrences of linked transactions

The condition of the enquiry parameter is checked within every loop in the main DAA for the option of occurrence validation. The event for which the occurrence is validated against is either in the past or in the future of the filtered object in the current loop. The occurrence value is assumed to be in minutes, and the variable named “data” is applied to the “takenOn” date for comparison.

Occurrence is a filtration mechanism through which the user decides to filter the results of the analytic parameter in the context of execution to a conditional satisfaction of a measurement’s existence within a certain period of time. Going back to the example provided in this section, the user might decide to filter the patients by those who had an open-heart surgery within the past 2 years. This kind of filtration empowers the analytical algorithm engine and provides the option of cascading different analytical enquiries from different users.

```
    } else {
     if (ep._EnquiryCondition.IsAnd) {
      last.setThis(filter);
     }
    }
   }

```

Figure 6.11: The case when no occurrence is specified

At the end of the loop, if the occurrence is not specified, the DAA filters the last performance record with the current filter as is.

As the main loop in this algorithm goes over all the conditions of all the parameters that are associated with the enquiry, main Identity filter is further developed to eventually serve as a filtration object to the main Identity of the algorithm, which would then be queried against the knowledge warehouse to generate the desired output.

```


id.PerformanceIdentities = new PerformanceIdentity[] {
 new PerformanceIdentity {
  _Performance = last
 }
};
```

Figure 6.12: Setting the last performance back to the main Identity

It is evident from the presented algorithm design that the user is able to construct complex queries that would normally be very difficult – if not impossible – to construct using classical analytical systems and would also be complex in nature even for technical specialists that are writing these queries directly against the underlying database.

6.5 Application Design

In line with the UML methodology presented in Chapter 2 and the resulting class diagrams that were presented in Chapter 5, this thesis produces a set of data-access classes in the C# programming language that handle all database access operations and hide the complexity of converting an object oriented query to a database query. Most databases use structure query language as their main querying language, but the logic injected in the data access classes is such that it adapts to the chosen underlying database. As such, the choice of a database for running the framework queries is not in scope of this writing, although for the sake of simplicity, this thesis adopts the community edition of the MySQL database engine. This enables deployment of the framework to any operating system, particularly Linux-based systems.

The building of a clinical analytical framework to host analytics for clinicians to execute based on customizable parameters is a complex objective that requires proper planning. The adoption of a solid methodology for engineering such a framework is key to producing rigid results. This thesis chooses a software engineering and development methodology that is flexible to allow for late-stage changes by the clients, but that is also manageable in-terms of resources and effort. The availability of several rapid software development frameworks and libraries helps reduce coding and testing efforts significantly but modularizing the software components and applying a loose-coupling paradigm that reduces dependency and increases reuse. The main design pattern followed throughout the development process is the MVC (Model Viewer Controller) software design pattern, which separates the data model from how this data model is represented in terms of controller objects, and then later viewed by presentation layers such as web pages and mobile views. The general architecture of this design pattern in any software system is shown in this figure:

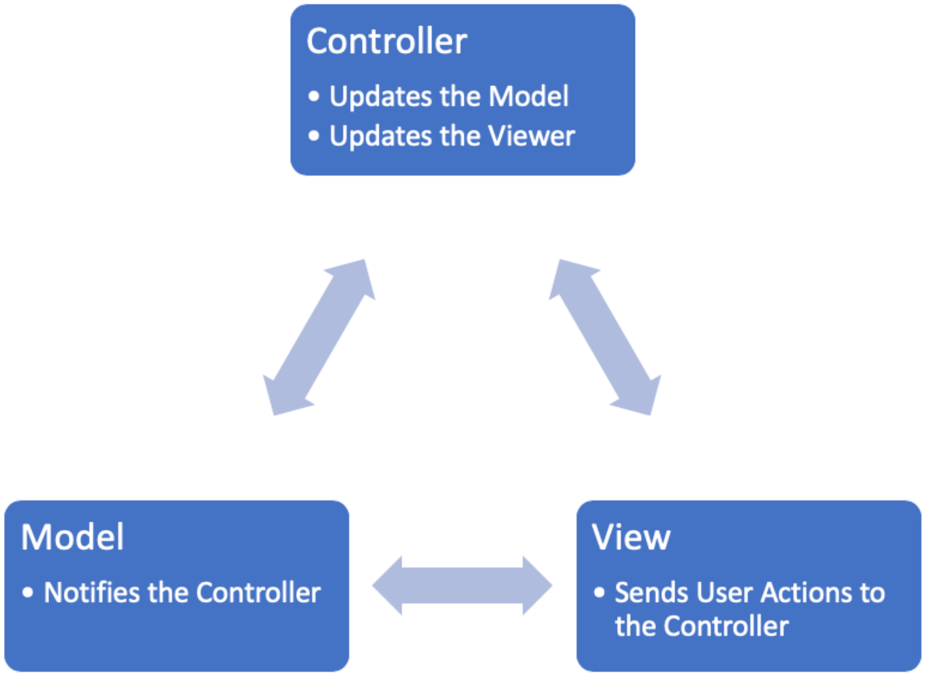

Figure 6.13: Model Viewer Controller software design pattern

6.6 The Framework System's Structure

Applying the methodology described in Chapter 2, the framework is which will host the parametrized analytical algorithms of the clinical intelligence healthcare data. The framework implementation approach will consist of four major stages:

- ***Class Diagrams:*** Creating the class diagrams that will define the data model of the entire framework. These will serve as the basis for creating the database schemas, the data-access-layer classes, and the basic data management forms in the system.
- ***Database Schema:*** Moving from the class diagrams, this thesis will create the corresponding database schema that will be the data warehouse. This thesis will define some additional database indexes for performance optimization, and This thesis will decide on the best choice of database table types.
- ***Data-Access-Layer Classes:*** The availability of data-access classes in the design architecture is critical to the success of the other phases of the system. These classes will coordinate the proper and easy delivery of data to and from the main data warehouse. They will act as a transitory data storage within the application context. The purpose of such classes is to isolate the data retrieval and storage activities from the business and process logic, and as such de-couples the process implementation from the underlying database choice.
- ***Web Data Methods:*** The creation of endpoints for communicating data over HTTP is a key feature of the framework, because this allows the creation of thin clients that produce and consume data to and from these web methods. The

approach here is like JSON and Web Services, with the adoption of a standard approach for building those methods.

- ***Model Containers:*** The framework is based on modern web-based technologies that can consume and produce data from HTTP endpoints. The model containers exist on both the server and client sides, allowing for great flexibility. The client can communicate back to the server through the HTTP data methods and therefore is capable of dynamically presenting the output and rendering a slick user interface.
- ***User interface:*** The user interface is built around a web framework that uses HTML5 technologies over a JavaScript framework for best user experiences. The framework allows also for mobile devices to connect, giving the analytical framework users maximum productivity options.

This thesis will now present the detailed design and implementation activities of the analytical framework. The activities are easily reproducible for people with a good level of technology and software engineering.

6.6.1 Application Server

The second phase of the framework design and development is the creation of an application server component that hosts the data model classes and data-access layer classes. The application server will deliver basic create, retrieve, update and delete operations (CRUD) and will also enable the client applications to execute advanced queries against the database through an embedded query algorithm within the data-access classes. The query engine will convert objects along with their dependencies into corresponding queries that will be executed against the data layer. The query engine uses a complex algorithm to eliminate the need for SQL-specific queries at the application level and at the same time improve efficiency of executed queries by using database-engine specific syntax and optimization hints.

The application server communicates with client applications an HTTP-based endpoint that receives the name of the operation and the request payload, executes the operation within the application context, and delivers back the result. The endpoint can be invoked in two modes: synchronous and asynchronous. In the synchronous mode, the response contains the operation result, while in the asynchronous mode, the caller is sent back in the response a reference to a running context at the server that can be polled for completion. A working example of both modes by performing the same operation are presented, which in this case will be a retrieval of the count of patients that had undergone surgery during the year 2016. As the patients are identities in the system belonging to the group “Patient” and is created as follows:

- Operation Request: The operation request body is the same for both synchronous and asynchronous calls.
- Synchronous call: The response to the request above in synchronous mode is simply the count distributed over the quarterly periods. The result is presented in the form of a JavaScript array.

The application server also serves as an orchestration or service-bus, which means that the application server executes the functionality requested on behalf of the client applications, reducing security risks and ensuring a single point of contact for the hosted data. Operations within the application server are implemented in separated logic units, like one-method classes in object-oriented programming models, and this design makes it easier for implementation isolation, where any failure or error would not affect the existing running operations. This unit structure makes deployment easier by introducing the notion of incremental builds, which adds functionality to the existing pool of operations supported by the application server. The application server acts as a translator, where it receives the payload in a text-based format such as XML, JSON or simple parameters in the URL, and produces an input set of parameters that are objects instantiated from data-access classes defined by the data model. The parameter objects are passed to the operation and the output of this execution is returned to the application server context where it is then converted to an appropriate output format, based on the request specifications. This three-stage execution mechanism allows for maximum functionality through a flexibility design that does not bind implementation to the execution environment.

6.6.2 Analytic Algorithms

The framework enables clinicians to build their own analytical algorithms in the form of computational units of logic that run within the context of the application server. The focus is on providing usable data sets that can be processed inside the algorithm logic without affecting the performance of other framework components. To achieve this, the application server provides the algorithms with contextual access to the data-access layer classes that server as pre-loaders that transfer data to an in-memory structure that can then be processed with other structures to achieve the desired results. The process is shown below:

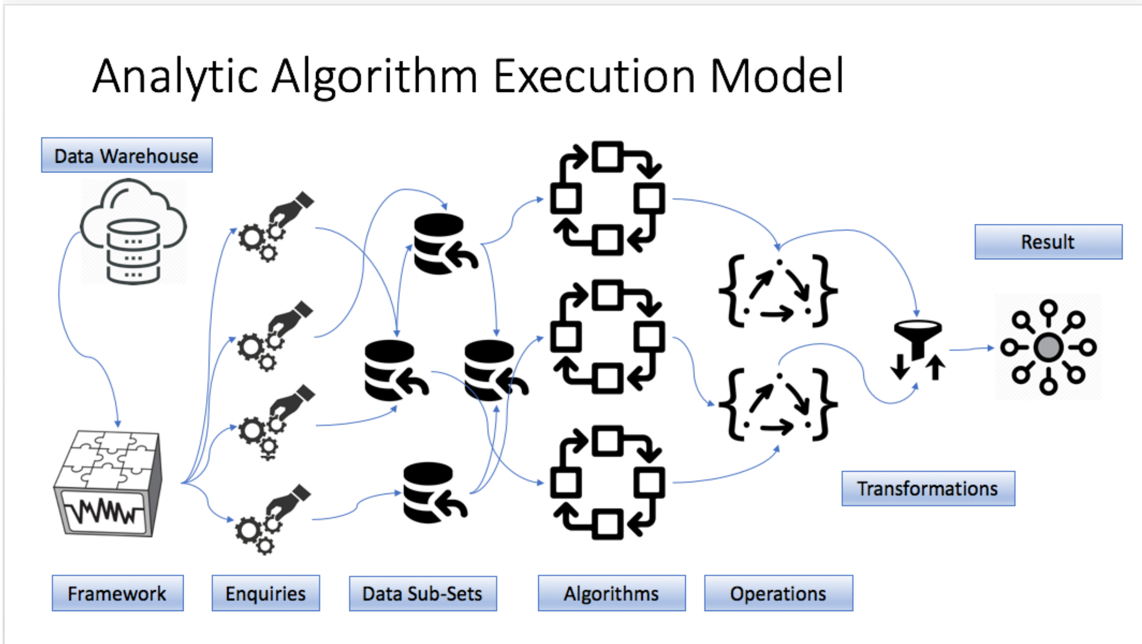

Figure 6.14: Analytic Algorithm Execution Model

The framework is connected directly to the data warehouse through the transparent data-access layer classes described above. The algorithms hosted within the framework context prepare requests for subsets of data that they need to perform their activities. The data sub-sets are defined as data structures that are populated by the execution of the enquiries against the data-access classes, resulting in collections of in-memory instances of the base classes. The analytic algorithm might require one or more such structures, and the assumption of the algorithm is that those structures are loaded before the algorithm execution begins. This allows clinicians to experiment with data that is not particularly up-to date and would reduce the load of algorithms testing on the application server data warehouse. Once the data sub-sets are loaded into memory the algorithm might be configured to perform a set of pre-defined or new operations on these data sub-sets, transforming their content to results that can be delivered back to the calling client. The algorithm consumes data and produces results within the context of the application server and is therefore in itself an independent entity. However, it could be possible to execute algorithms from within the running context of other algorithms, allowing for algorithm chaining. The figure below shows the result of re-using analytic algorithms within the application server context.

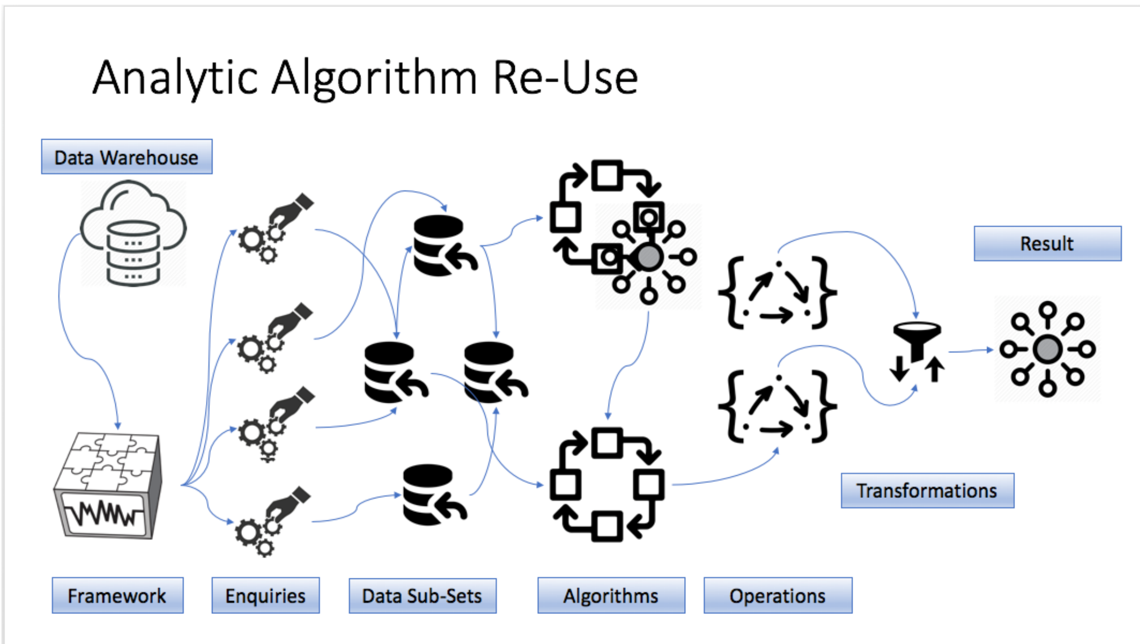

Figure 6.15: Analytic Algorithm Re-Use

The re-use of existing analytic algorithms is possible because algorithms consume structures of data as data-subsets and produce results in the form of data structures, which are either filtered subsets of input data or aggregated results. In both cases, the data is usable as an input data-set to other algorithms. The chain process can go on to include multiple input-results and might result in multiple output datasets.

6.6.3 Application Clients

The application server design makes it easy for clients to connect and execute server-side functionality through secure and flexible endpoints over standard communication protocols such as HTTP. Such protocols and their associated ports are considered secure for outbound connections by almost every firewall in any network environment and are the most popular protocols for communication for most mobile application providers. For the web client implementation, the modern approach of creating units of visual content that represent pages in a large application container is followed. This idea is like the one used in the application server model and allows to build the pages that produce and consume data with such flexibility that can be loaded dynamically into any HTML5-compliant container. These pages have a template visual design that is populated with data coming from the application server as the page is loaded into the container and is currently the most popular client-side content presentation mechanism and is seen in many application frameworks such as Bootstrap, Angular-S, and Sencha, and has been recently introduced into operating systems such as Windows and Linux and are most commonly referred to as widgets. The figure below shows the conceptual structure of widgets.

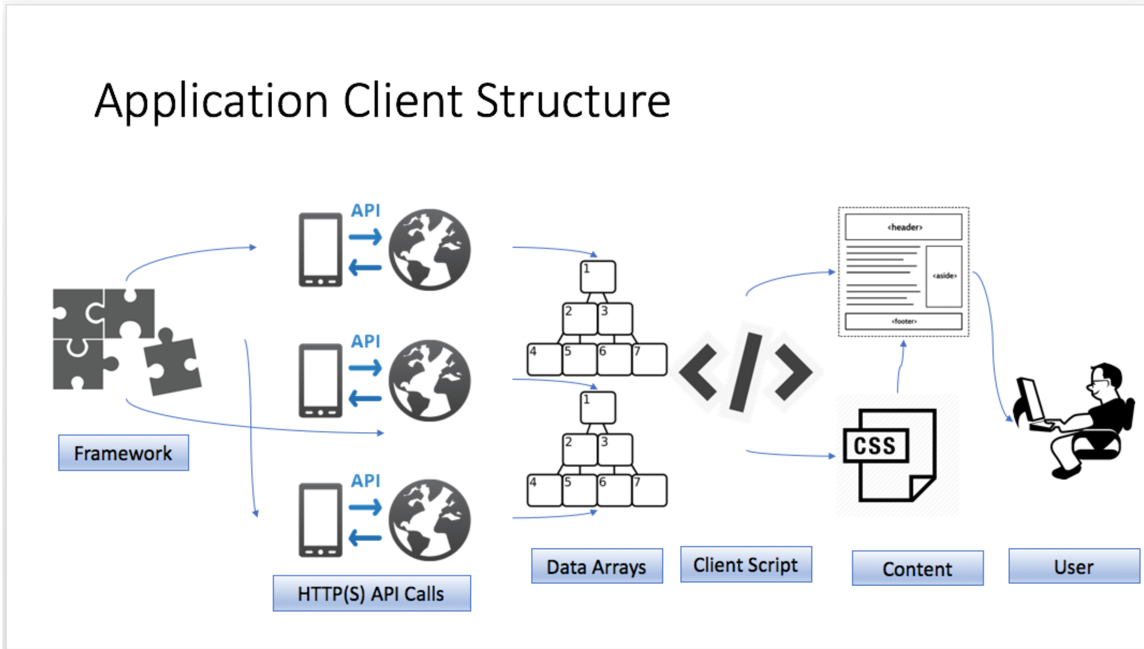

Figure 6.16: Application Client Structure

These widgets are units of visual content, and are composed of a client-script, content data-holders, and one or more templates. Within the context of the application that is delivering them, the widget would be loaded by the application framework within a visual context such as an HTML page, a windows form, or a mobile page. The widget loads the data sources that it needs to present its content asynchronously, usually in the form of JavaScript objects, and once all data structures have been loaded the application framework delivers this data to the templates of this page, and these in term are executed against the data objects to generate presentable content, which is in tern styled using the CSS styles. This process is context-independent, so the widget does not need to be aware of what context or application is loading it, but rather is implemented with no dependencies. This design decision does limit the functionality of the widget, but also makes it application independent, thus reducing content maintainability and creating a single user-experience across all platforms. To further improve maintainability and deployment efficiency, the widgets are stored within the application server and delivered through similar endpoints to the executing application contexts.

6.6.4 Visualization Engine

The widget concept is also used to create visualization contents, where templates are replaced with charts and gages that allow for data visualization using different chart types. Most of the available charting toolkits for HTML5 are array-driven, and thus the concept for data loading that is shown in this chapter is compatible with these chart widgets. The widget data might need to be transformed to a structure that is readable by the chart object and might also require some grouping based on the chart type, but the overall process is similar.

6.6.5 API endpoints

The application server is designed with interface-capabilities in mind and is therefore able to interact with other systems using different communication protocols and deliver content in multiple formats. The usage of application server communication endpoints is subject to proper design of the methods that are being processed. As previously explained in Section 6.6.1, the system defines a set of methods that act as containers for loading data from the data warehouse through data-access classes. The main requirement for such methods to deliver content back to clients through endpoint calls is that they define a return structure per content-type. For example, if the content-type is JSON, then the method should format its output into a JSON object. The application server context provided to the method defines the calling context, and thus the method could respond accordingly. While the method implementation itself is not affected by the content-type required, it is the responsibility of the implementation to handle multiple content-types. The framework provides built-in functionality that helps convert objects to different content-types, reducing the overhead of repetitive coding. The framework also uses templates to inject code back to the receiving client.

The API uses an XML-based payload to delivery request content to the application server. The choice of the payload structure was based on the completeness of XML in terms of its ability to represent complex structures, in addition to its portability and cross-language interchangeability.

6.6.6 ETL implementation

The implementation of the ETL phase is now possible based on the foundations of application server and web applications communication design. The ETL data model had already been presented in Section 5.5 and allows for a data field mapping transformation activity that could include several transformation steps. As this is part of the framework application design, it is important to provide the steps that can be taken leading to the administration of a successful data loading activity. The implementation will be explained by using a simple example to prove the concept, with a sample loading of a data-set from a regular admission department in a local hospital. The data is provided in the form of a spreadsheet application file, but that could also be any other format that provides tabular data. As such data sources are frequently provided for updates, the first step is to define a data map that would be applied to all future data. As explained in previous chapters, the data loaded from this file would be linked to a batch record that is created during the load operation.

6.7 Application Scenario

In order to properly use the framework system as presented in this chapter, an application scenario is presented in this section that would help implement the framework system as part of a monthly status report creation process in a hospital environment. The status report is composed of several sections and parts that are usually created by different departments and units. The output is compiled and presented on a monthly basis to the top management.

An outline of the usage was divided into the following phases:

- Phase 1: The data models from the different data sources needed to generate the hospital monthly status report would be defined. The department managers would contribute to the model definition.
- Phase 2: The data warehouse would be imported from the previous month. This is the same data that was collected, aggregated and merged into the status report of last month. It is assumed that the status report of last month is the most accurate one.
- Phase 3: The analytics of each department's contribution to the status report would be created. The users would be assisted to develop their analytics, which would also serve as a training.
- Phase 4: The monthly status report would be created from the departmental analytical algorithms. The report would then be compares to the previous month's manual report.
- Phase 5: The status report for the current month would be created the following month. The results would then be checked and validated against the current month's manual method.

The phases above are usually the same with any other application scenario and would not be specific to the application scenario in scope. The first phase requires that the data models be identified based on the existing understanding of the users for the data that they usually use. An example of this is the weekly admissions data, where the admissions process would report a list of patients that were admitted during a week in time. This would usually be a tabular list, where the different columns represent data that is of importance to the analysis of this output. The process owner would then know how the columns are related, and whether two columns represent information about the same identity or not. For example, the process manager – and probably anyone else in the healthcare organization – would know that the patient admission number and the patient name are both columns that identify the patient with a code and a name value, and would therefore belong to the same identity, namely the "Patient". In other columns the values that are calculated would also be identified. An example would be the "time to bed", where the information system might record the difference in time between the end of the admission process and the time the patient reached the bed in the room of admission. Values such as this would help the process managers answer questions from top management regarding the efficiency of the processes in execution. For a successful loading of the data from a data source, the data owners should define three sets of fields: the identity fields, the value fields and the date field.

The second phase could be either manual or automated, based on the accessibility permissions that are put in place. For some systems the data access is possible to define, and other systems could connect and retrieve data from its exposed data sets or reports. For other systems the data is not accessible through endpoints or interfaces and would need to be exported manually and then imported into the framework system. The format of the exported tabular data can be any of the commonly used formats such as comma-separated-values (csv) lists. Regardless of the import mechanism, the pre-defined data map from phase one is applied and would then generate the performance entries in the underlying database.

For the third phase, the analytics would be defined and tested based on the imported data from phase two. It is usually the case that the clinicians and users import a sample test data set which can be then tested manually against the same data in another statistical or spread-sheet tool. The important part of this phase is that the clinicians and users feel confident that the results generated by the framework system are the same as would be generated manually on a small data set. This confidence in the generated results would then clear the confusion that would later be faced with larger data sets that would be difficult to process with classical tools.

The fourth phase of the application scenario involves an integration of the results from different analytics into a single view or dashboard that would be later presented in the fifth phase of the process. The results could be assembled from the different analytic outputs using a third-party platform that would present the results in a clearer way. In the execution of the application scenario, a decision-support portal was used to serve the output to the top management, due to the fact that top-management personnel wanted to get the results without connecting back to the hospital's internal and private network, and for that sake the output of the analytical executions would need to be exported and then imported to the online portal.

6.8 Summary, conclusion and discussions

This chapter outlines the implementation details of the proposed framework from the previous chapters. The chapter follows the methodologies of software development that were explained in Chapter 2. The chapter develops based on the analysis from Chapter 5 and starts with an argument that the framework is feasible in terms of implementation as a software system. The chapter then explains the Dynamic Analytical Algorithm that is the core of the analytical framework and provides an example to further clarify the steps in the algorithm. The detailed technical and structural design of the framework is then explained in the next sections, covering all modules of the framework.

The chapter makes it clear that non-technical analytical users in a healthcare environment could build and maintain a structured knowledgebase from different data sources and systems, together with a dynamic analytical engine that brings the data closer to the consuming user, be that a decision-support officer or a senior clinician.

Chapter 7

Evaluation of the Framework System

7.1 Introduction

Evaluation of information systems is a challenging task that can be approached with different methods and techniques, and for each a different result may be reached. The domain of healthcare poses a greater challenge since it is a hybrid environment of different and sometimes conflicting information systems that occupy the same landscape with multiple common data sets and users. This chapter evaluates the framework decision-support system that was developed in the previous chapters and puts it to work in a healthcare environment that depicts a similar picture to most healthcare centres. The chapter starts by reviewing existing literature to best understand the criteria, challenges and practices of information system evaluation techniques and methodologies, in a special focused approach and following the guidelines of Chapter 2. The chapter then continues with the actual evaluation process, involving the healthcare users and clinicians in the implementation of the framework decision-support system and getting their feedback through a questionnaire that is later analysed to reach a quantitative evaluation which is merged with a qualitative evaluation from different meetings and feedback. The chapter concludes with a presentation of the results and recommendations for future improvements.

7.2 Evaluation Literature Review

The argument in [25] is based on the weaknesses of evaluation methods that only focus on asking the questions that identify the people, the processes, the timelines and the procedures of implementing the IS. In a sense, the work criticises evaluations that are only based on technical evaluations and is rather focusing on the implementation and adoption of an IS as a phenomenon that is to be studies from the early stages of the idea adoption to the final stages of outcome evaluations. The qualitative rather than quantitative approach is emphasised in [25], and therefore many subjective ideas are promoted as leading to a better evaluation. The authors argue that measurements such as quantity and size are the in scope when the evaluation is performed in a quantitative approach, but the subjective approach that they discuss is focused on the investigations of the surrounding aspects of an implementation, why the implementation was decided, and how the implementation is performed, in a way that meets the decision-makers' demands. The work suggests a mixture of both methodologies to perform a successful evaluation. [25] points out two deficiencies in the current methodologies for IS evaluation. The first is that such evaluations focus on the identification of the subject of IS implementation, rather than the reasons why the implementation was undertaken and the decisions that lead to its adoption. The second deficiency is the lack of a framework for IS evaluation, and therefore there is no agreement on the domains of evaluation, and the evaluation approaches.

The work in [37] analyses results from questionnaires that were given to users of various electronic health systems in both private and public sectors. The authors argue that the accuracy of the questionnaire results was higher than would be for a typical one-system questionnaire, because the questions were asked not based on system functions, but rather on how the system improved the overall performance of the setting in which it was operated. The users were addressed with questions regarding usability, performance, impact on running processes, patient safety issues, and adaptability to system functions. The results served as an evaluation guideline to various EHR systems in use today, focusing on the major difference between public and private sector information systems. [37] reported that physicians in private sectors were in general satisfied with what the information system produced, whereas their public-sector counterparts had a negative evaluation of their own systems. The work continued to present varying usability and adoption rate results that support this dissection of the user base. The authors reveal that in some situations, users of public-sector information systems have identified situations or incidents when the information system's lack of functionality has led to considerable effect on patient safety. The work identifies factors that are most important to users and are therefore important aspects of any future healthcare information system evaluation. The factors include performance, usability, downtime, bad design, relation of the system to the existing processes and its ability to adapt to their changes, and finally the adoption rate of a system in a healthcare setting.

The complexity of the evaluation process in a healthcare environment was discussed in [87], where different evaluation methods were contrasted in an attempt to derive a best-result evaluation methodology. The healthcare setting involves not only the technical aspects of a software system utilization, but also the human and social aspects as well. There is a vast difference in the way a software system is utilized in a healthcare setting, where the priority is the quality of the patient engagement rather than the accuracy of the deliverables. [87] recognizes that in addition to the challenges of a technical evaluation in a healthcare environment, the evaluation process itself involves other different objects to be evaluated.

The continuous monitoring of quality feedback from system users and stakeholders is found to be very important in [14]. The work emphasises the importance of having a quality framework for evaluation, where the users and the system evaluation officers can efficiently and continuously monitor the changes in system acceptance and usability.

The study of [47] was used as a guideline to further identify an efficient method for the evaluation of the software framework presented in this thesis. The work in [47] performs a classification of several methods and techniques from the literature of software evaluation, resulting in a clear classification of three categories of evaluation techniques and methods. The work finds that predictive methods most common, followed by opinion-based methods, and lastly behaviour-based methods. In that regard, the evaluation method that is adopted in this thesis and that was presented in Chapter 2 is a clear combination of all three categories and would therefore serve the evaluation purposes best.

To further provide validity and credibility to the evaluation of the framework system detailed in this thesis, various evaluation methodologies would be required to evaluated in a systematic and measurable way that gives a guidance into the evaluation activity that this chapter implements. The work in [71] performed a similar comparative study leading to a measurable set of criteria that help identify the most suitable evaluation methodology from the literature studies, based on dimensions that the work defines. [71] represents every evaluation method as a point in a six-dimensional space of axes representing the value of the method in terms of “criterion, evaluation technique, form of evaluation, secondary participants, level of evaluation, and relativeness of evaluation”. By clustering the methods under study using the distance function in this six-dimensional coordinate system, the work in [32] was able to identify seven mostly used styles of software evaluation. A demonstrative approach is found to be among the most used styles, and other styles include metric and simulation absolute evaluations. The work in [32] shows clearly that a combination of evaluation styles, techniques and methods is most efficient when a software information system is in scope of evaluation. The healthcare environment poses challenges in every domain, and for information management, knowledge management and decision-support the challenges are bigger.

A similar study is conducted in [41], but in a way that uses communication through meeting with evaluators of an information system to identify a list of unique problems arising from the usability experiences. The work determines the severity of each problem from this list by assigning a value from 0 to 4, indicating the degree of impact that problem on the users of the system, the frequency at which the problem occurs, and the whether the problem is lasting or not. [41] compares the results for 2 methods, the cognitive walkthrough and the heuristic evaluation. The work finds that while a cognitive walkthrough method for software evaluation identifies usability problems better, the heuristic evaluation method detects problem that eventually dissatisfy the users.

This chapter builds on the works reviewed by adopting a software evaluation method that gives clinicians and decision-support officers a comprehensive and quality-oriented approach to evaluation of the decision-support framework introduced in this thesis. The chapter starts with describing the steps taken to deliver a questionnaire to the users of this system, and then proceeds to evaluate the results of this questionnaire by examining a focus group’s response and how it relates to the global response set.

7.3 Questionnaire: Feedback from the users

Based on the software information systems' evaluation methods' literature review from the previous section and following the methodologies' review from Chapter 2, this section discusses the implementation of the evaluation methodology adopted in this thesis. The methodology execution is divided into four stages:

- Meetings: introductory and evaluator meetings with clinical stakeholders and top medical and healthcare management officers, and then group meetings with departmental and organizational sections to collect usability expectations and social impact analysis.
- Questionnaire: a list of problems, fears, and expectations are compiled following the meetings, and a categorized list of questions are stated that focus on different aspects of the meeting outcomes.
- Results analysis: an analysis is performed to understand the distribution of answers and to assert the usefulness of the software framework, based on a focus-group evaluation approach where the answers from key users of different sections are analysed. Further meetings are conducted with these key users to get detailed verbal feedback.
- Findings presentation: the findings of this evaluation are presented to the top management and stakeholders to inform them of the overall status of the software information system usage and how it has changed the state of the organization.

Following the methodology plan above, the execution started by scheduling meetings with two top management executives at a local hospital. The meetings were conducted in person at the location of the officers. The officers were the hospital general manager and the hospital owner. The meetings were scheduled independent of each other and did not involve any other personnel. The first meeting was with the hospital general manager, who informed of the necessity of a healthcare quality of service and knowledge maintenance mechanism that keeps all departments aware of the impact of their operations on other departments and on the global healthcare services of the hospital. The general manager also informed of the need to eliminate the cumbersome efforts of generating a monthly general report. The hospital managers are asked on a monthly basis to provide data related to their department's activities in order to consolidate this data into a monthly hospital general status report. This will include medical, operational and financial data in the form of metrics that guide the hospital management into making key decisions for future hospital improvements, including services focusing, new hires, equipment investment, research funding and global spending for the purposes of improving the general hospital standings and ranking. A copy of the report for a recent month was given as an example, with details on how it is created.

The second meeting was with the hospital owner. The meeting followed the meeting with the general manager and in this meeting the report and concerns of the general manager were discussed in a similar setting. The hospital owner informed that the report is very important to move the hospital forward, but another equally important requirement that was not possible yet is the ability to directly create this report without the need for any departmental involvement. The hospital owner was afraid that the data that is submitted by the department manager could be somehow manipulated to expose a status that is incorrect. The hospital owner understood that data could come from several systems that are currently in operation at the hospital, and that new systems could come that would be either privately used by clinical services or globally by the entire hospital. The owner stated that no matter how complex, numerous or detailed the systems are, it was prohibitively impossible to manually collect, compile and prepare the data needed to generate the monthly report. The owner had also investigated several business intelligence systems that would help in this regard but was not willing to hire a technical team to manage this process, in addition to the main concern that any internal staff member from the technical department could manipulate the data or intervene with the outcome.

Following both stakeholder meetings, other meetings were held with department managers and process owners. The first of those meetings was with the medical management, represented by the chief medical doctor and the nursing manager. In this meeting the general directions of the top management were discusses, both the monthly status report and the data security and safety concerns. The medical and nursing management informed that in the medical cases, information is stored in the hospital management system after the patient case is handled. They informed that the medical analysis is a process that relies heavily on accumulated treatment plan outcomes, and that most doctors lack such information, either due to access difficulties when they encounter the patient, or due to the complexity of preparing this information prior to the patient treatment. The most notable concern was medical errors, which might occur when a treatment plan is put in place that is not the most optimal from previous encounters. The medical director informed that due to the large volume of patient encounters, the hospital has collectively enough information to treat most new patient cases based on previous knowledge but is not efficiently reducing the heal ratio due to lack of knowledge transparency and analysis.

The second middle management meeting was with the pharmacy department, represented by the head pharmacist. The pharmacy is a 24/7 service that receives electronic patient orders from the hospital system, which are signed by the treating doctor. The pharmacy delivers the required medication or replaces non-existing stock with another generic drug alternative. The pharmacy delivery mechanism is based on pick-up and return, where the drugs are ordered electronically, picked-up by the nurse from the pharmacy, and then returned both electronically and physically to the pharmacy. The key concerns for the pharmacy were the large volume of drug returns due to overestimation of treatment requirements, the large volume of consumables that are requested without the ability to negotiate the needs, the urgent supply shortages due to seasonal viruses for example, and limited stock supplies from suppliers that receive orders for immediate delivery.

The next meeting was with the admission department manager, who expressed deep concerns regarding the monthly status report activities. The manager explained how the report data is collected, where different reports are exported and then aggregated using a spreadsheet tool. Following that the data is mixed to fulfil the management's data collection requirements. The department manager stated that when the data is in question, the process needs to be reversed so that the source of the error in the original aggregation is found. Since the top management compares the admission data with patient heal ratios and re-admissions with same symptoms, the admission department requires that doctors clearly state the diagnosis of the patient before admission.

Following the meeting with the admissions management, a meeting with the quality management was held, where the quality manager estimated that 20% of the monthly effort of the quality department is spent collecting data from different departments to consolidate them and create a monthly quality of service report, similar to the one for the top management. The QoS report shows the quality department and other top management staff the global quality status of the hospital per service and per department, focusing on failures and incidents that were reported through the incident management system in place. The incident management system is not linked to the hospital system, and for that matter data aggregation and preparation is important.

The last meeting was with the ancillary services managers, which included managers from departments such as the laboratory, the radiology, and the operations services. The services have in common the interest of delivering the fastest and most cost-efficient services to the nursing floors. The main focus in their discussions were two key performance indicators that all of the services maintain on a monthly basis. The first is the average time-to-service for every patient. This is the average sum of all patient service time, from the time the order is received to the time the result is delivered or the outcome is achieved. The second key performance indicator was the cost-per-patient, where the large sum of costs needed to deliver any service is divided into parts that are not always easy to compute or aggregate. The costs could include human and material costs, in addition to time loss and resource allocation errors.

Following the meetings with the top and middle management officers the framework developed in this thesis was introduced in a series of presentations that focused on how each problem that the departments had would be solved. The top management decided that it would use the system across the entire hospital for two months, following the application scenario presented in Chapter 6, with a focus on benchmarking the results against the manual application scenario that was previously in place.

The plan was implemented for two months in the middle of the year, where the hospital load was low, and the staff were more available to contribute to this effort. Following the implementation of this plan, a questionnaire was created that would be used to evaluate the system implementation, usage and outcomes. The questionnaire was developed with the guidance of the work reviewed in section 1 of this chapter, and from the meetings that have been conducted before the system adoption plan was implemented. The users were allowed a few months following the implementation plan to use the system further. In addition, new and more complex analytical tasks were conducted by clinicians and nursing management to get information related to treatment plan success and recovery patterns. The questionnaire was presented at the end of the year, and included members from the following departments and groups:

Group	Members	Focus Data Sets	System Utilization
Top Management	2	Everything	Moderate
Pharmacy	3	Drug Sales, Consumption	Frequent
Administration	2	Invoices, Occupation Rate	Moderate
Nursing	5	Diagnoses, Orders, Results, Occupation Rate, Operatory Data	Frequent
Admissions	2	Occupation Rate, Admission distributions, Invoices	Moderate
Clinicians	4	Diagnoses, Admissions, Orders, Results, Demographics, Occupation Rate	Frequent

Table 7.1: Questionnaire Distributions

All questions in the questionnaire were multiple-choice questions with one allowed answer per question. Most answers were based on the Likert scale and the answers to every question were: 1- Strongly disagree, 2- Disagree, 3- Neither agree nor disagree, 4-

Agree, 5- Strongly agree. At the beginning of the questionnaire is a section where the users answer questions related to the nature of their work, the degrees they have, and other information related to the proper assessment of the users' experience with the system. The remaining questions were statements that needed to be confirmed or rejected by the user and would therefore have a definitive answer regardless of the involvement of the user with the system feature or conflict in question. At the end of the questionnaire the answers were checked to make sure that no question has more than one answer. Users who did not want to answer a particular question were asked to choose the answer "3- Neither agree nor disagree", which would then correspond to a neutral response to the question.

The questionnaire was divided into five sections, the first one related to the user's evaluation and the others with a focus on a particular conflict type. The following table lists the sections of the questionnaire, with a description of each section and the questions within it. The conflict types were compiled based on the outcomes of the meetings and the implementation plan explained in this chapter.

Section	**Questions**	**Significance**	**Conflict Type**
Assessment	4	3 – Low	None.
Usability	3	2 – Medium	Time-consuming manual reports preparation
Data Access	4	1 – High	Multiple data sources and systems
Analytics	3	2 – Medium	Information based on extracted data is not always simple to generate
Output	3	3 – Low	Representation of findings is sometimes difficult, due to lack of visualization
Reusability	2	2 – Medium	Time is wasted re-doing the same tasks for generating the monthly report by different departments
Collaboration	3	2 – Medium	Outcome from one analytical task is usually not available for others to use.

Table 7.2: Questionnaire Sections

The choice of questionnaire sections was based on the characteristics of the software system and the success factors' categories from different literature works that have been reviewed in Chapter 3, with a clear focus on software evaluation work from the evaluation methodologies discussed in Chapter 2, particularly the classification of success factors for information system implementations from [84]. The factors that were found of most impact to a successful evaluation of a software system were focused around the ability of the users to assess the software system objectively, their ability to use this software system, either assisted or unassisted, the measure of their expectations compared to what the system actually delivers, their ability to build on their or other users' results, and how they work together as a team to achieve better results. The questionnaire sections were also aligned with the conflict points identified following the different meetings with the hospital staff and based on the analysis performed in Chapter 5. The questions were thus compiled and are shown in the below table:

N	O	P
Question	Section	Description
Question 1	Assessment	What is your Highest Degree?
Question 2	Assessment	What age group do you belong to?
Question 3	Assessment	How would you describe your computer skills?
Question 4	Assessment	How many years of experience do you have in healthcare?
Question 5	Usability	I was able to use the system without much help
Question 6	Usability	The system showed me the information that I needed
Question 7	Usability	I was able to complete my task without problems
Question 8	Data Access	The systems' data in my organization is available for processing
Question 9	Data Access	I was able to load my own data without problems
Question 10	Data Access	The data that I loaded was consistent with the systems that I use
Question 11	Data Access	I was able to recognize the categorizations from the systems that I use
Question 12	Analytics	The system gave me results that summarized the base data
Question 13	Analytics	The queries that I created represented valid knowledge
Question 14	Analytics	I was able to combine data from different systems
Question 15	Output	The results were clearly presented
Question 16	Output	The results were easily understood by my management
Question 17	Output	The results could be compared with previous results
Question 18	Reusability	My effort was reduced by existing simplifications in the system
Question 19	Reusability	I was able to easily create my own customizations
Question 20	Collaboration	I was able to benefit from queries that my colleages used
Question 21	Collaboration	I was able to use my own queries more than once
Question 22	Collaboration	I was able to share my work with team members

Figure 7.1: Evaluation Questionnaire sections and questions

The questionnaire was answered by the users through one-on-one meetings that were conducted over a period of one month, with as little time gaps between the meetings as possible. The users were asked at the end of the meeting to disclose the information discussed or shown in the questionnaire, for the sake of fair analysis of the results. Users were urged to answer all questions, but in some cases, users preferred not to answer. All questions were answered in a limited and equal time-frame among users. The users were informed before they started the questionnaire that some sections were of more importance to them based on their work requirements that others, and in that the users were indirectly informed that the sections of most importance should be carefully thought-of. The responses were kept completely anonymous and no reference to the user or department that is filling the questionnaire is found.

7.4 Results of the Questionnaire

Based on the literature review in the first part of this chapter and following the general evaluation methodology guidelines presented in Chapter 2, this section analyses the results from the questionnaire responses based on quantitative and qualitative analysis techniques. In the quantitative part of the evaluation, the results are aggregated and segmented based user groupings from the Assessment section answers. This segmentation allows better understanding of the response in addition to giving meaningful descriptions of the averages and other data aggregation functions applied. Then the responses are analysed per question, where each question is aggregated based on the responses of another question. This process helps show the relationship between the answers to questions that might exhibit a logical correlation. The last part is the qualitative analysis of the results, where the acceptance of the system is determined based on user feedback and productivity improvements.

The users who evaluated the system using this questionnaire were using the framework system for achieving their monthly tasks of creating segments of data aggregations as part of the monthly status report to the stakeholders and top management. The way they used the framework system is detailed in Appendix B.

A correlation analysis was performed using SPSS's correlation tables and charting functionality to detect the relationships between different question combinations in the questionnaire answers. To simplify the correlation identification, only two questions at a time were correlated together.

Basis	Correlation	Sections	Table	Figure
Question 1	Question 2	Assessment	Table 7.4	Figure 7.2
Question 3	Question 4	Assessment	Table 7.5	Figure 7.3
Question 3	Question 5	Assessment, Usability	Table 7.6	Figure 7.4
Question 5	Question 7	Usability	Table 7.7	Figure 7.5
Question 7	Question 6	Usability	Table 7.8	Figure 7.6
Question 3	Question 9	Assessment, Data Access	Table 7.9	Figure 7.7
Question 10	Question 11	Data Access	Table 7.10	Figure 7.8
Question 12	Question 10	Analytics, Data Access	Table 7.11	Figure 7.9
Question 3	Question 12	Assessment, Analytics	Table 7.12	Figure 7.10
Question 3	Question 13	Assessment, Analytics	Table 7.13	Figure 7.11
Question 13	Question 14	Analytics	Table 7.14	Figure 7.12
Question 14	Question 15	Analytics, Output	Table 7.15	Figure 7.13
Question 14	Question 16	Analytics, Output	Table 7.16	Figure 7.14
Question 16	Question 10	Output, Data Access	Table 7.17	Figure 7.15
Question 18	Question 20	Reusability, Collaboration	Table 7.18	Figure 7.16
Question 3	Question 20	Assessment, Collaboration	Table 7.19	Figure 7.17
Question 3	Question 21	Assessment, Collaboration	Table 7.20	Figure 7.18
Question 3	Question 22	Assessment, Collaboration	Table 7.21	Figure 7.19

Table 7.3: Questionnaire correlations mapping

The above table shows the correlation combinations chosen for analysis based on two selection criteria. The first selection criteria are that both questions should be in the same section whenever possible, and the second is that the questions have a common relationship identified from the previous sections of this chapter. Following the meetings with the hospital users and management, questionnaire questions that reflect a common evaluation outcome were marked for correlation.

The responses are aggregated first by assessment groups. This step is based on the fact that the first part of the questionnaire includes an assessment section that would help classify the users into groups. In the assessment section four questions were presented to the user related to personal information. The user is asked to choose the highest achieved degree, the age-group that he or she belongs to, the level of their computer skills, and the experience years he or she has on the job. Every question in the assessment section has five answers, from which the user can choose only one. The following tables show the distribution of users based on the responses from the four questions:

Q1/Q2	18-20	21-35	36-50	51-60	61+
None	1	0	0	0	0
High School	0	2	1	1	2
Certification	0	1	1	2	0
Bachelor	0	0	1	1	0
Masters+	1	2	2	0	0

Table 7.4: Degree to Age responses distribution

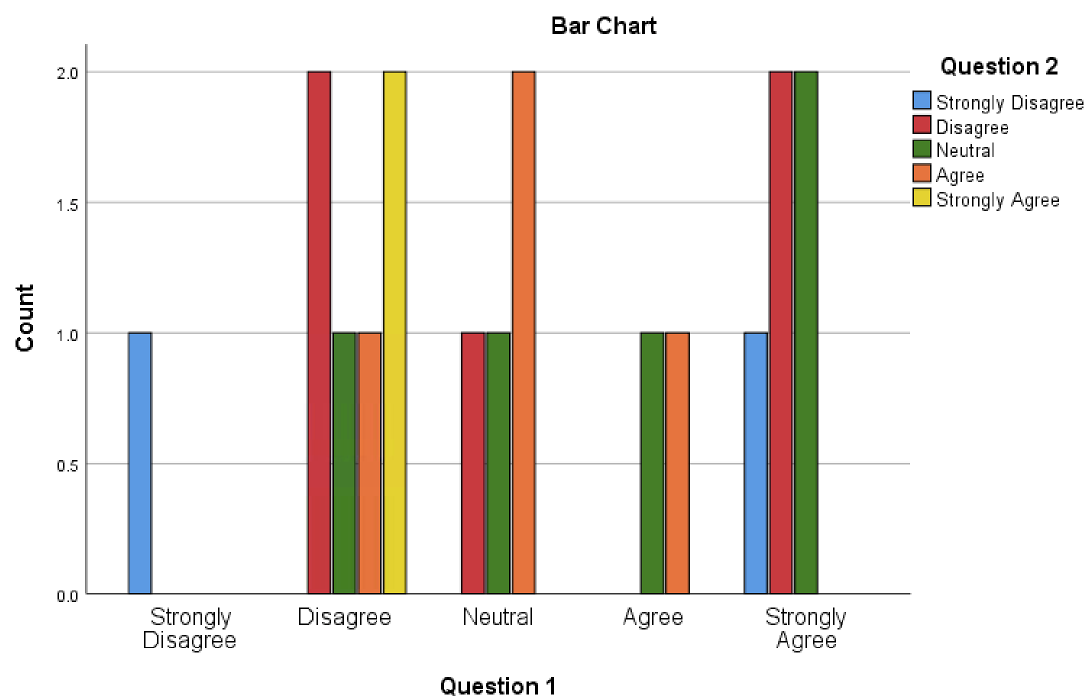

Figure 7.2: Degree to Age responses distribution

Q3/Q4	0-5	6-10	11-15	16-20	21+
None	2	0	1	2	0
Basic	0	0	1	0	0
Moderate	0	1	2	0	0
Good	1	0	0	2	0
Expert	1	0	0	4	1

Table 7.5: Computer skills to Job years responses distribution

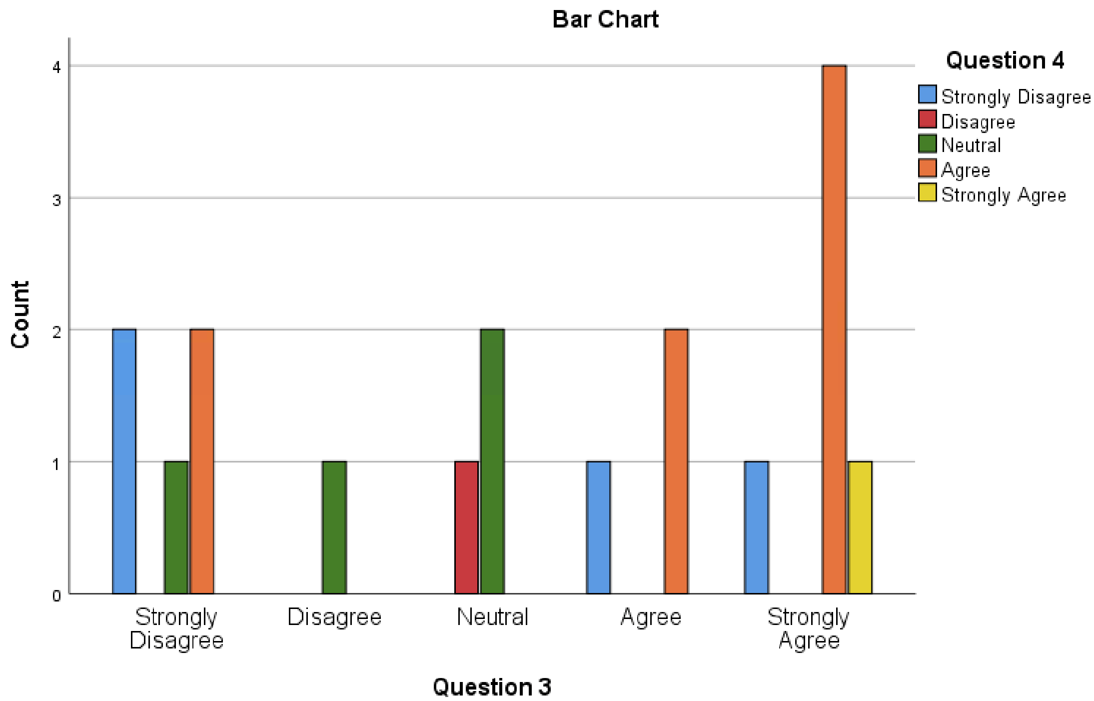

Figure 7.3: Computer skills to Job years responses distribution

The assessment section shows a relatively large number of employees with expert computer skills and years on the job with a moderate concentration of users around the middle ranges in job years and having moderate computer skills. However, a few employees do have no computer skills and have some years on the job. The results are expected in a population from the healthcare setting. As a matter of fact, many healthcare organizations would have employees who are in a reporting position to the top management who might not have enough computer skills but rather have a long experience in their actual job.

The usability section of the questionnaire asks the users to validate, based on the Likert acceptance scale, three questions related to how easy it was for them to use the system:

- **Question 5:** I was able to use the system without much help
- **Question 6:** The system showed me the information that I needed
- **Question 7:** I was able to complete my task without problems

The general nature of the questions related to what exactly the task that needs to be completed and what information the user needed was intentionally meant to be differently conceived by the users based on their personal understanding of the information and tasks. The following shows a distribution of computer skills based on the ability to use the system without help, as per the responses from question 5:

Q3/Q5	1	2	3	4	5
None	1	0	2	1	1
Basic	0	0	1	0	0
Moderate	0	0	1	2	0
Good	0	0	1	2	0
Expert	0	1	2	1	2

Table 7.6: Computer skills to Unassisted system usage distribution

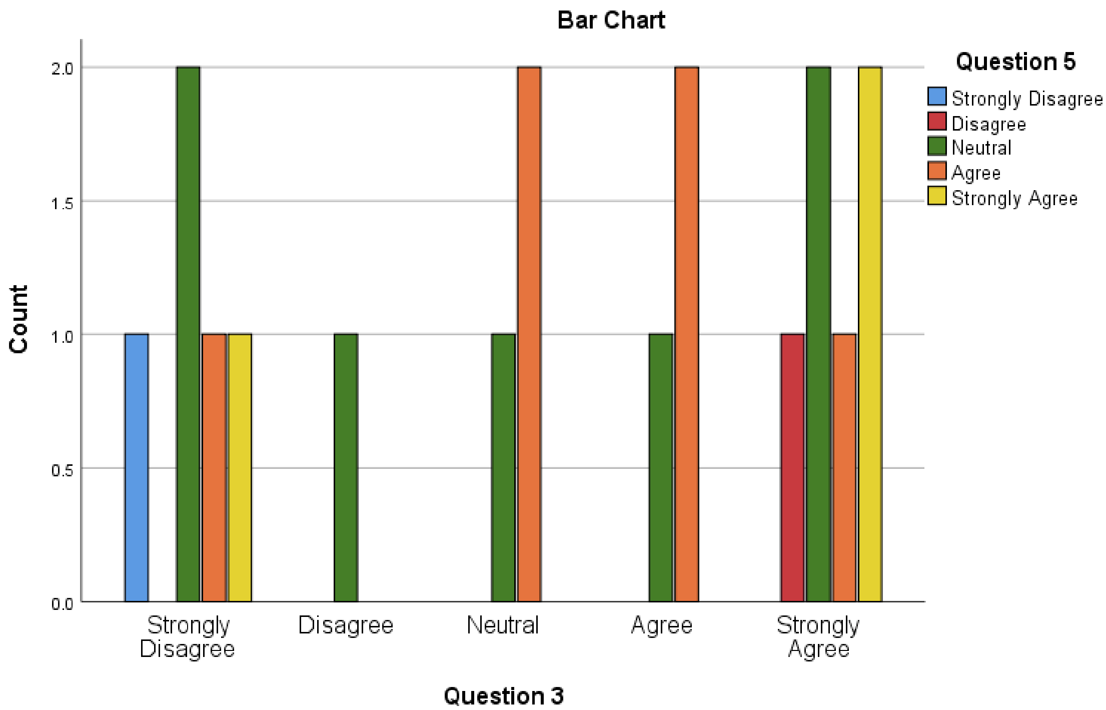

Figure 7.4: Computer skills to Unassisted system usage distribution

The table shows that most users were able to use the system on their own, with a good number of users with no to moderate computer skills who could do this on their own. The number of self-learning users with moderate, good or expert skills is also apparent in the distribution.

Another distribution then would show the system's utilization by the users, where the users who were responses to the users' abilities to use the system on their own is compared to their ability to complete their intended tasks. The following table shows this distribution:

Q5/Q7	1	2	3	4	5
1	0	1	0	0	0
2	0	0	1	0	0
3	1	1	0	3	2
4	0	1	2	0	3
5	0	0	0	1	2

Table 7.7: Distribution of responses based on Unassisted usage and Task completion

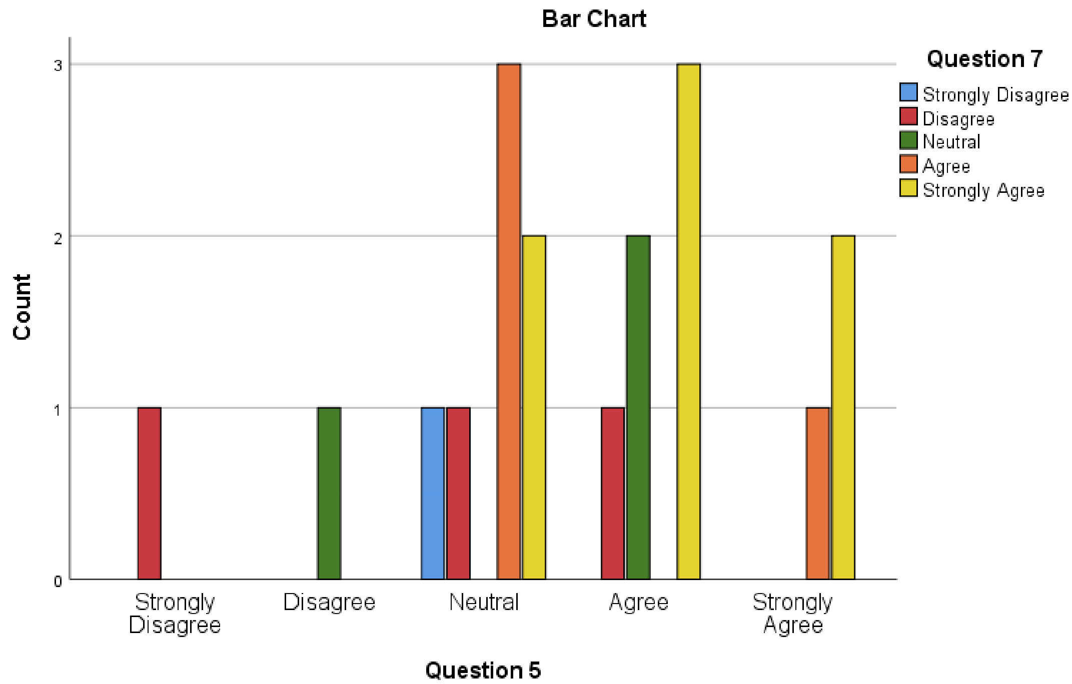

Figure 7.5: Distribution of responses based on Unassisted usage and Task completion

The distribution table shows clearly that most users who could use the system on their own were able to do so to complete the tasks that they had. This conclusion can then be augmented with another distribution, showing if users were able to complete their tasks with the framework's available information:

Q7/Q6	1	2	3	4	5
1	0	0	0	1	0
2	0	1	0	1	1
3	0	1	0	0	2
4	1	0	1	1	1
5	2	1	1	3	0

Table 7.8: Distribution of responses based on complete tasks and information availability

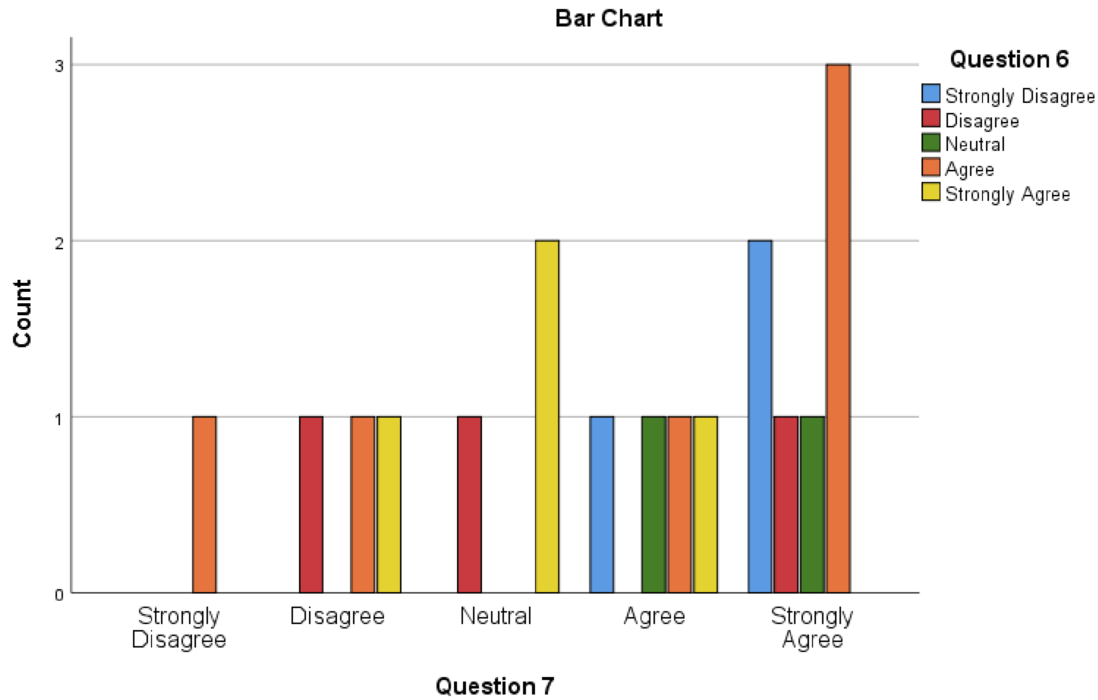

Figure 7.6: Distribution of responses based on complete tasks and information availability

In the distribution above the availability of information in the system was not evidently shown to be the reason why users were able to complete their tasks. When combined with the findings in Table 7.6, the results can be interpreted as representative of the ability of the users with good computer skills to find their way through the system and request data that is not already there in the system. This conclusion would need further confirmation from the responses that were given in the “Data Access” section, where users were asked to present their acceptance of the following four statements:

- **Question 8:** The systems' data in my organization is available for processing
- **Question 9:** I was able to load my own data without problems
- **Question 10:** The data that I loaded was consistent with the systems that I use
- **Question 11:** I was able to recognize the categorizations from the systems that I use

To check whether the conclusion that skilled users were able to get the missing information, the responses from the computer skills question are compared to the responses from Question 9 to check how users with access to the system data could load missing information in the system:

Q3/Q9	1	2	3	4	5
1	1	1	0	2	1
2	0	0	1	0	0
3	0	1	1	0	1
4	0	0	2	1	0
5	1	0	2	1	2

Table 7.9: Distribution of computer skills and easy data loading

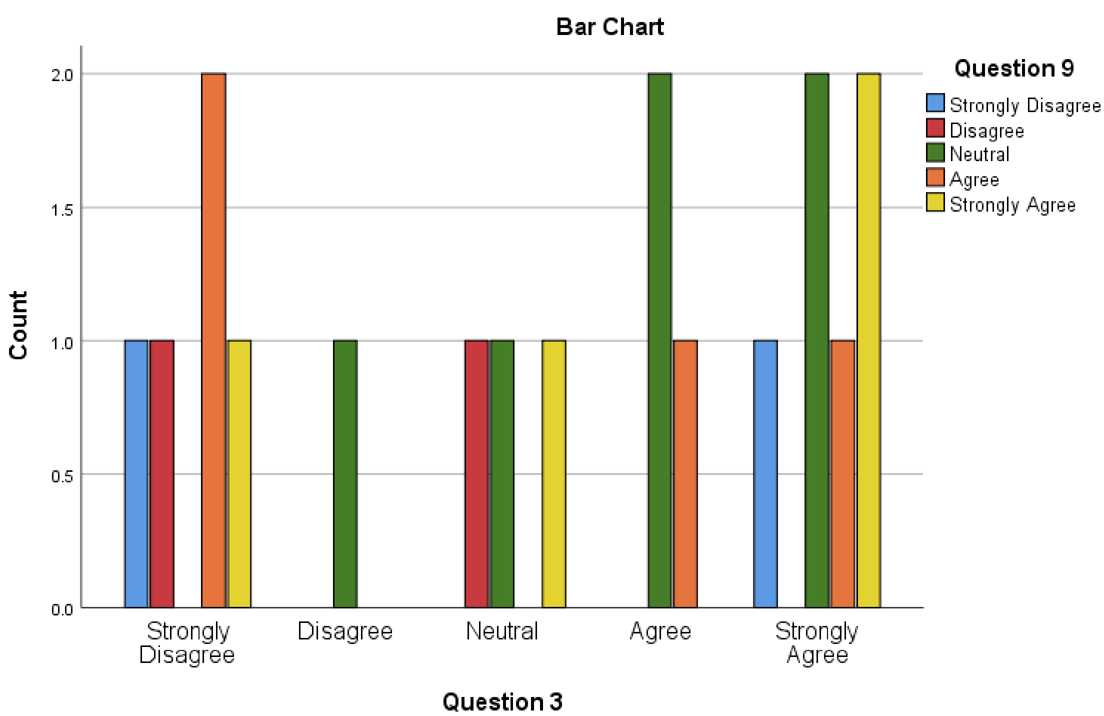

Figure 7.7: Distribution of computer skills and easy data loading

The distribution shows clearly that most users faced very little problems loading their own data. This confirms that when users with moderate computer skills used the system, they were able to use the system unassisted and were able to complete their tasks either by finding the existing data in the system or being able to load this data from the external systems.

The quality of the loaded data in the system is also investigated based on the responses of the users. In the "Data Access" section users were asked to confirm whether the loaded data was consistent with the existing data in other systems, and whether the data ended up properly categorized in the frameworks warehouse. For this purpose, the responses from questions 10 and 11 need to be analysed together. The following table shows this distribution:

Q10/Q11	1	2	3	4	5
1	1	1	0	0	2
2	1	0	0	1	0
3	0	0	1	2	1
4	1	0	2	0	0
5	2	0	2	0	1

Table 7.10: Distribution of responses based on the quality of the imported data

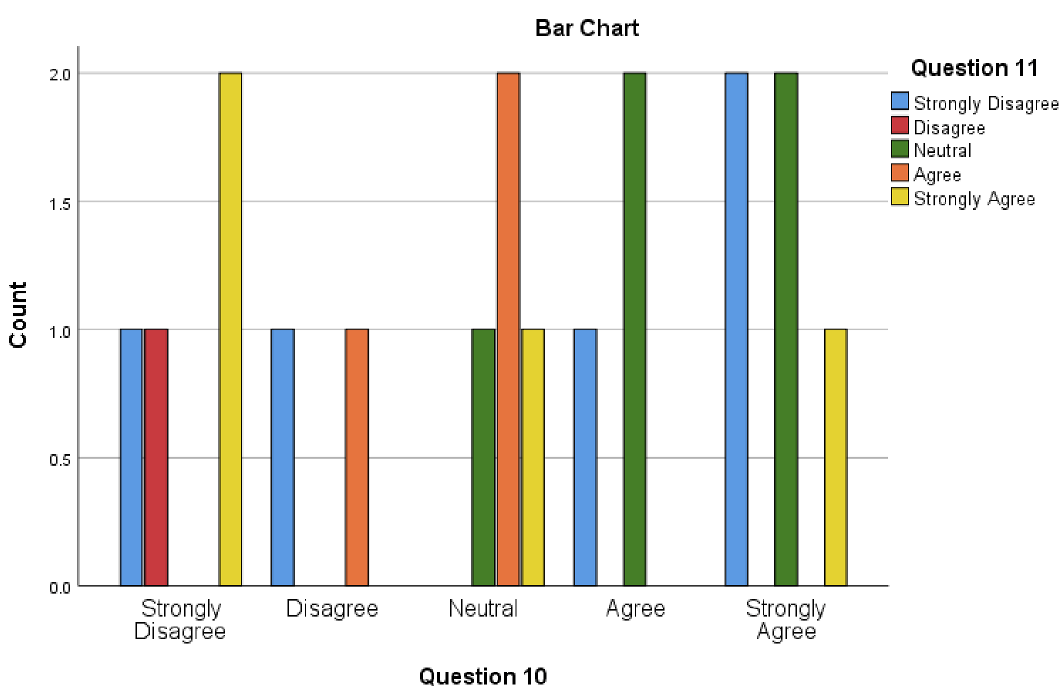

Figure 7.8: Distribution of responses based on the quality of the imported data

The distribution shows an almost even balance between responses to the users' ability to identify the categories based on how consistent the loaded data was with the systems that the users use. This means that users were able to load data from the existing systems, but do not have an agreement on the quality of this data.

The fourth section of the questionnaire asks the users to assess three statements:

- **Question 12**: The system gave me results that summarized the base data
- **Question 13**: The queries that I created represented valid knowledge
- **Question 14**: I was able to combine data from different systems

The importance of the statements in this section is that they establish a combined assessment of how the users were able to achieve outcomes with the system that would otherwise require more effort if done manually or with other existing solutions. The users' responses in this section would also be checked against the distributions analysed in the previous questionnaire sections, giving a better understanding of which user group produced the better results. The responses from Questions 12 would provide further understanding of how the users were able to move forward with their analytical tasks, as it is expected that users be able to properly and easily summarize their data sets. The responses are compared to the responses from Question 10 related to the consistency of the imported data from the underlying systems:

Q12/Q10	1	2	3	4	5
1	0	0	0	0	3
2	1	1	1	2	0
3	2	1	0	0	0
4	1	0	1	1	1
5	0	0	2	0	1

Table 7.11: The distribution of responses related to analytical work and data quality

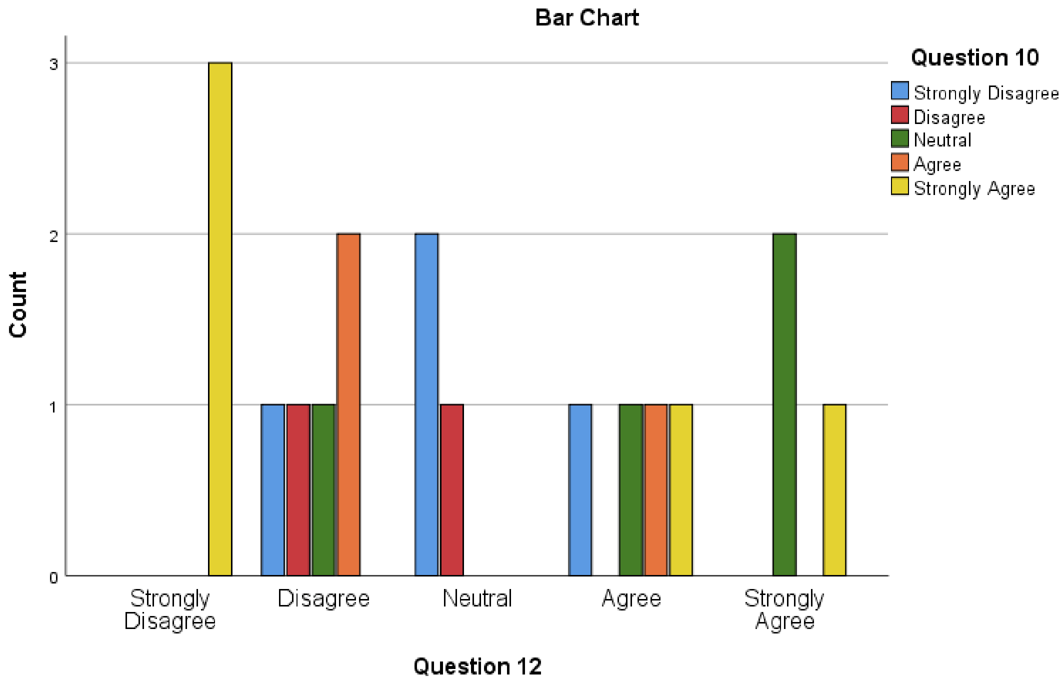

Figure 7.9: The distribution of responses related to analytical work and data quality

The responses show an acceptable confirmation of the ability of users to perform summarizations on their data. The confirmation is not completely clear from which user group it is coming, so another analysis of the distribution based on the computer skills of the users would need to be done. The following table shows the distribution of users' computer skills with their ability to perform data summarizations from Question 12:

Q3/Q12	1	2	3	4	5
1	1	0	1	1	2
2	1	0	0	0	0
3	0	2	0	1	0
4	1	1	1	0	0
5	0	2	1	2	1

Table 7.12: Distribution of responses based on computer skills and summary of data

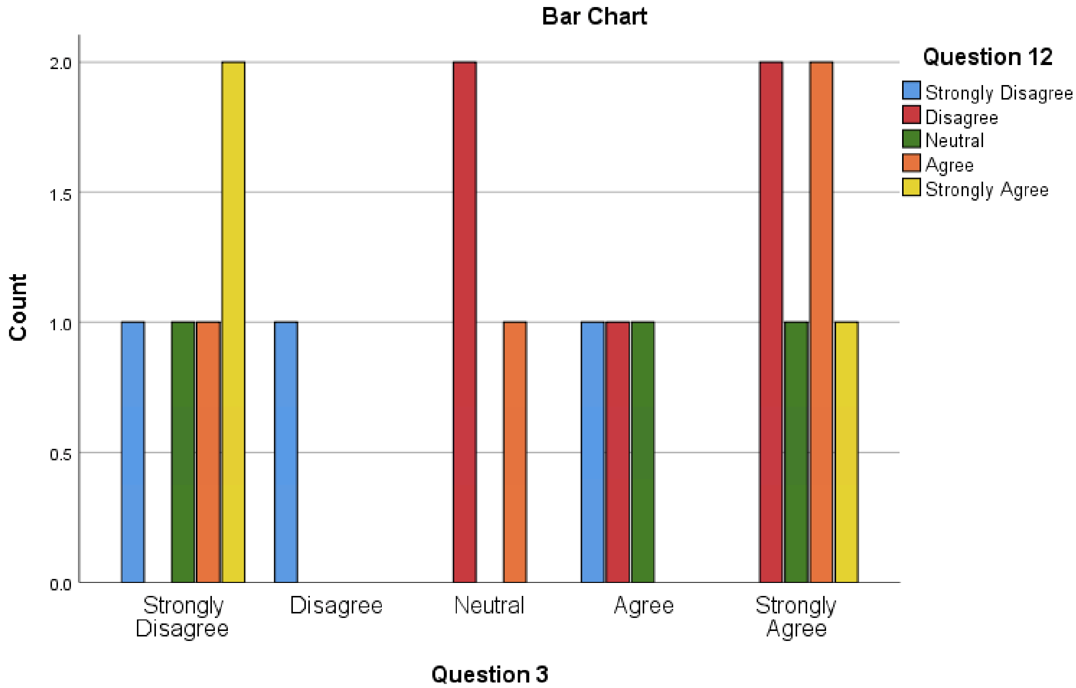

Figure 7.10: Distribution of responses based on computer skills and summary of data

The distribution above clearly shows the ability of most skilled users to perform summarizations of their data in the system framework under evaluation. The distribution also confirms that users with moderate to low computer skills were also able to summarize their data. Few users with good computer skills were not able to summarize their data, so the distribution in the table above should also be compared to responses from question 13, where users were asked to confirm whether the queries that they created resulted in valid knowledge. The information helps understand how skilled users were able to generate a working information set that they could use in their analysis:

Q3/Q13	1	2	3	4	5
1	1	0	1	3	0
2	0	1	0	0	0
3	0	1	0	2	0
4	0	1	1	1	0
5	0	1	2	2	1

Table 7.13: Distribution of valid knowledge output with computer skills

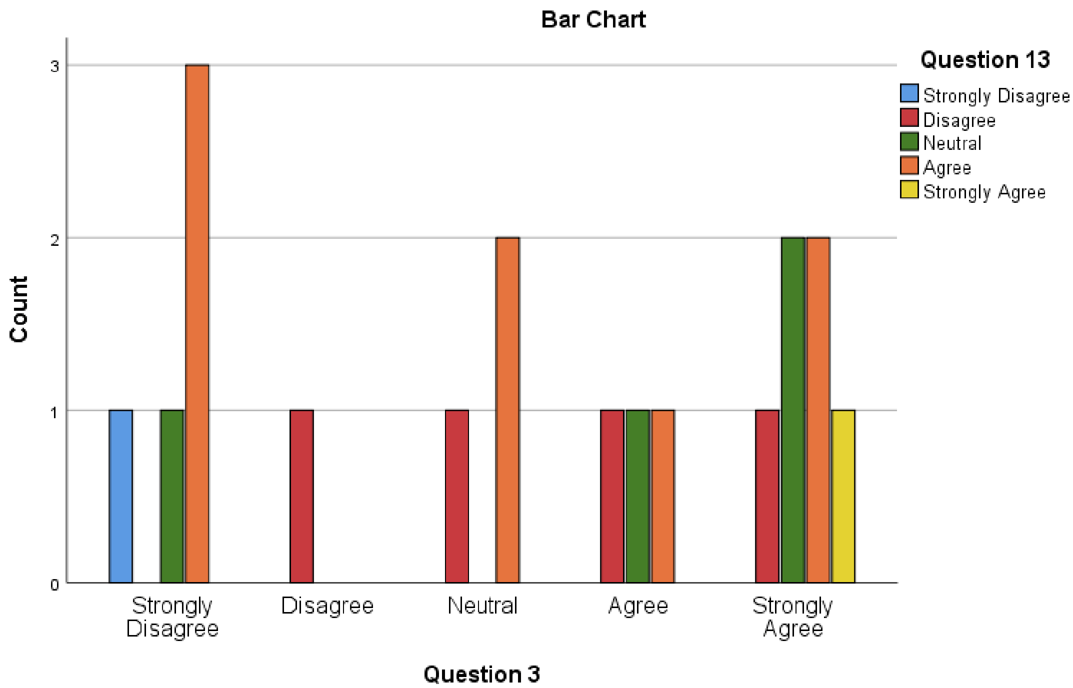

Figure 7.11: Distribution of valid knowledge output with computer skills

The above table shows that the majority of users with good computer skills were able to generate information from their analysis that they perceived as valid. This also shows that some users with little to no computer skills were also able to create useful knowledge out of the analysis, which they might have been able to do in an assisted manner. The distribution clearly indicates the ability of users to use the system's output knowledge for further analysis in their assigned tasks. Another distribution would also show how those users were able to combine knowledge into new knowledge or fact that they used to further improve their deliverables. This is shown by comparing the results from question 14 with the results from question 13:

Q13/Q14	1	2	3	4	5
1	0	0	0	1	0
2	1	1	1	1	0
3	0	1	0	1	2
4	1	1	4	1	1

5	0	0	1	0	0

Table 7.14: Distribution of valid knowledge to data combination responses

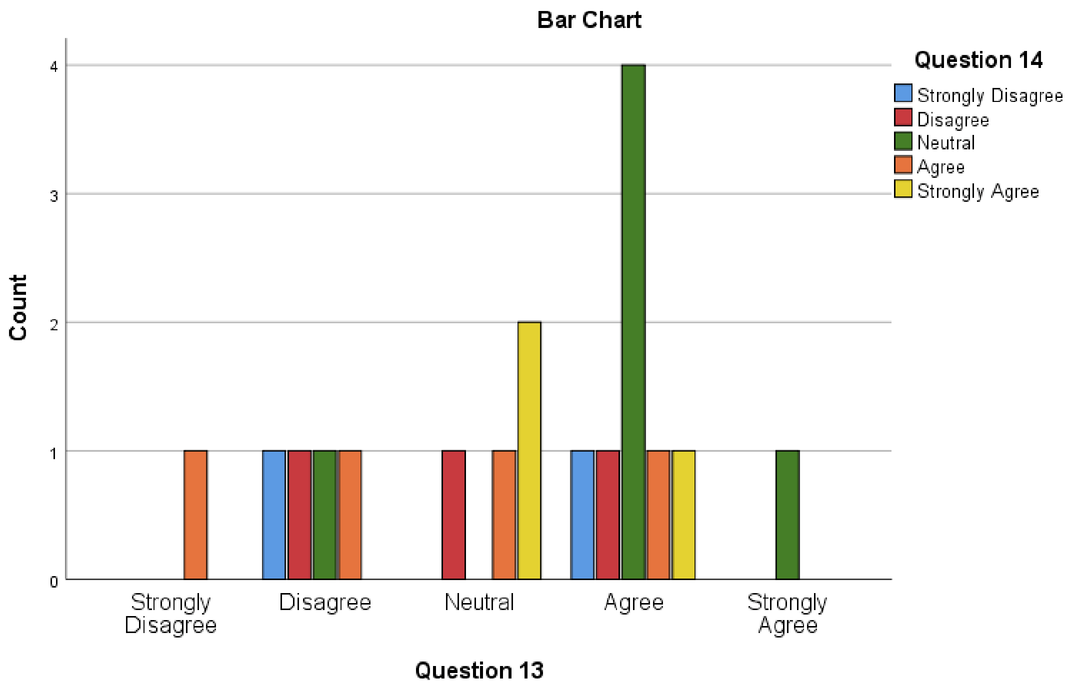

Figure 7.12: Distribution of valid knowledge to data combination responses

The comparison shows clearly that users were mostly able to combine data that they analyzed and generate new knowledge based on it. The statement that "*most users were able to generate new knowledge from existing analysis that they performed using the framework system*" can be asserted with a high degree of accuracy, while the statement that "*users were able to summarize existing data that they loaded from existing systems*" could be asserted with a moderate degree of accuracy.

As the users created new knowledge with an acceptable degree of acceptance, the questionnaire responses would then be used to validate statements related to the fifth section of the users' satisfaction with the framework system, namely the "Output" section. In this section the users are given three statements to validate:

- **Question 15**: The results were clearly presented
- **Question 16**: The results were easily understood by my management
- **Question 17**: The results could be compared with previous results

Comparing the responses from Question 15 with the users' ability to combine data from different underlying systems (Question 14) would show if the combined data could be presented properly using the framework system:

Q14/Q15	1	2	3	4	5
1	0	0	1	0	1
2	1	1	1	0	0
3	1	0	2	3	0
4	1	2	0	1	0
5	0	1	0	2	0

Table 7.15: Distribution of combined data with data presentation

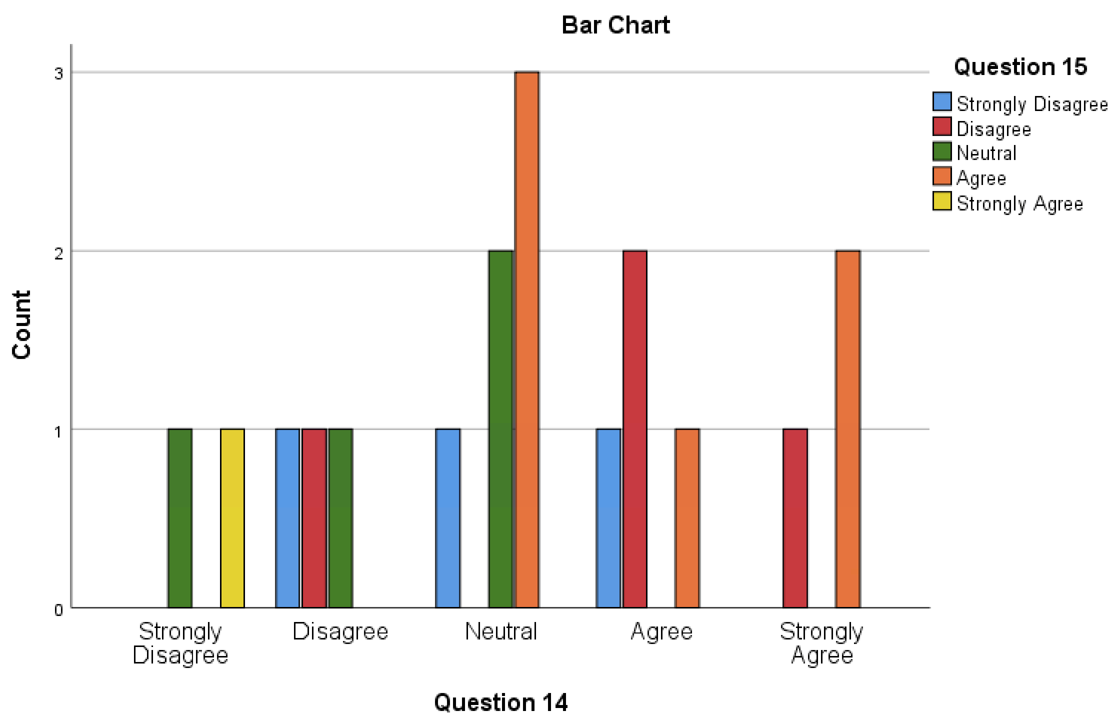

Figure 7.13: Distribution of combined data with data presentation

The results show that the majority of users who were able to combine data from multiple systems were also able to present this data in the system outputs. Combining these results with another comparison with the responses from Question 16 would show if the combined data was reported as clearly understood output by the management:

Q14/Q16	1	2	3	4	5
1	0	1	0	0	1
2	0	1	1	0	1
3	0	2	1	1	2
4	1	0	0	2	1
5	1	2	0	0	0

Table 7.16: Distribution of combined data with management understanding

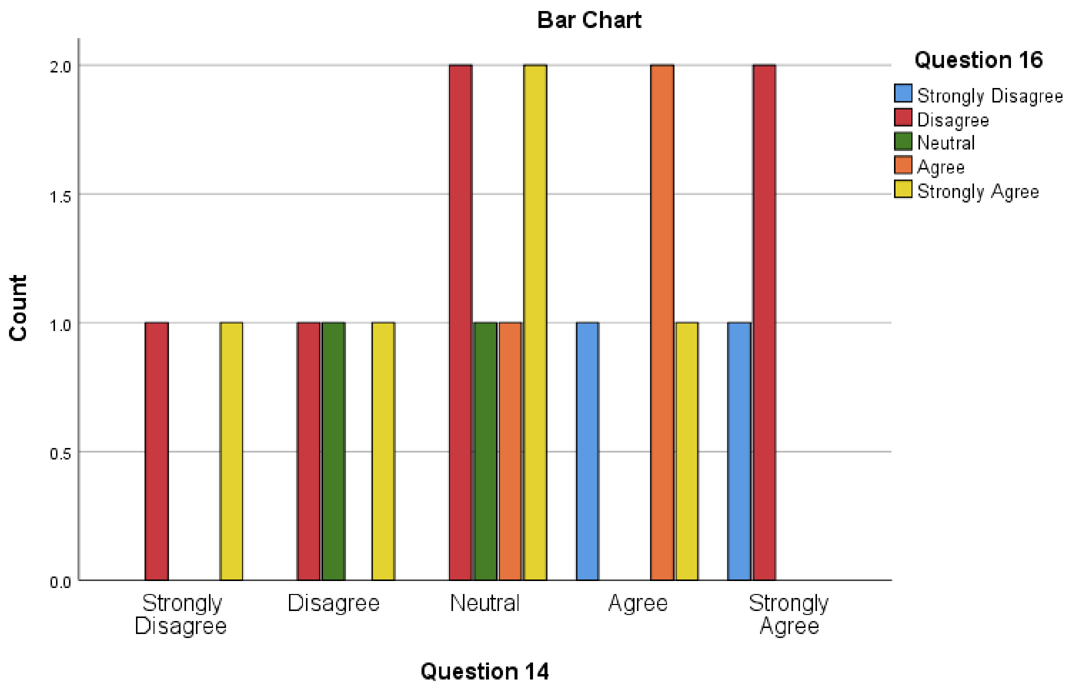

Figure 7.14: Distribution of combined data with management understanding

This shows an even distribution of responses related to the top management's acceptance of the output generated by the system as related to the users' ability to combine data from different underlying healthcare systems. The results reflect a certain relentlessness of the top management towards the framework system's output but would also need some further analysis to check whether the top management acceptance is related to the quality of the data that is initially loaded into the system. A comparison is needed with the responses from Question 10 that is related to the proper loading of data that is consistent with the underlying systems' data. The following table shows the responses from Question 16 with the responses from Question 10:

Q16/Q10	1	2	3	4	5
1	1	1	1	1	0
2	0	1	0	0	1
3	0	2	0	1	1
4	0	0	1	1	1
5	1	2	0	0	2

Table 7.17: Comparison of data quality with top management's acceptance of results

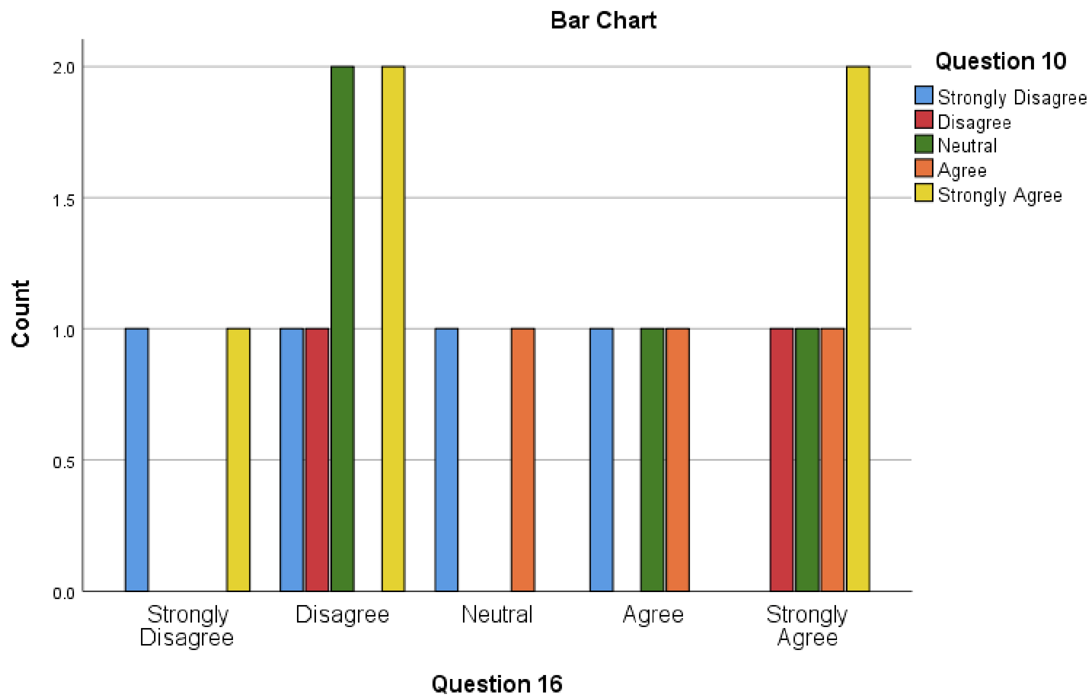

Figure 7.15: Comparison of data quality with top management's acceptance of results

The results show that most output is accepted by top management but is related to an even distribution of quality data. This means that while the data that is initially loaded into the system did not completely satisfy the users, they were able to work with it to a point that was later acceptable by top management, a result that shows the ability of the users to transform the quality of input data into a reasonable output that was accepted by the top management.

The last two sections of the questionnaire included general questions related to the reusability and collaboration of the output data. These statements were presented in the following questions:

- **Question 18**: My effort was reduced by existing simplifications in the system
- **Question 19**: I was able to easily create my own customizations
- **Question 20**: I was able to benefit from queries that my colleagues used
- **Question 21**: I was able to use my own queries more than once
- **Question 22**: I was able to share my work with team members

The two sections' statements could be understood better when compared with each other. The framework's existing simplifications (Question 18) would be compared to the system's ability to enable to the user to benefit from existing objects that other users created (Question 20). The following table shows this comparison:

Q18/Q20	1	2	3	4	5
1	1	3	0	0	1
2	2	1	0	2	1
3	2	1	1	0	1
4	0	0	0	1	1
5	1	3	0	0	1

Table 7.18: Distribution of system simplifications based on reusability of other objects

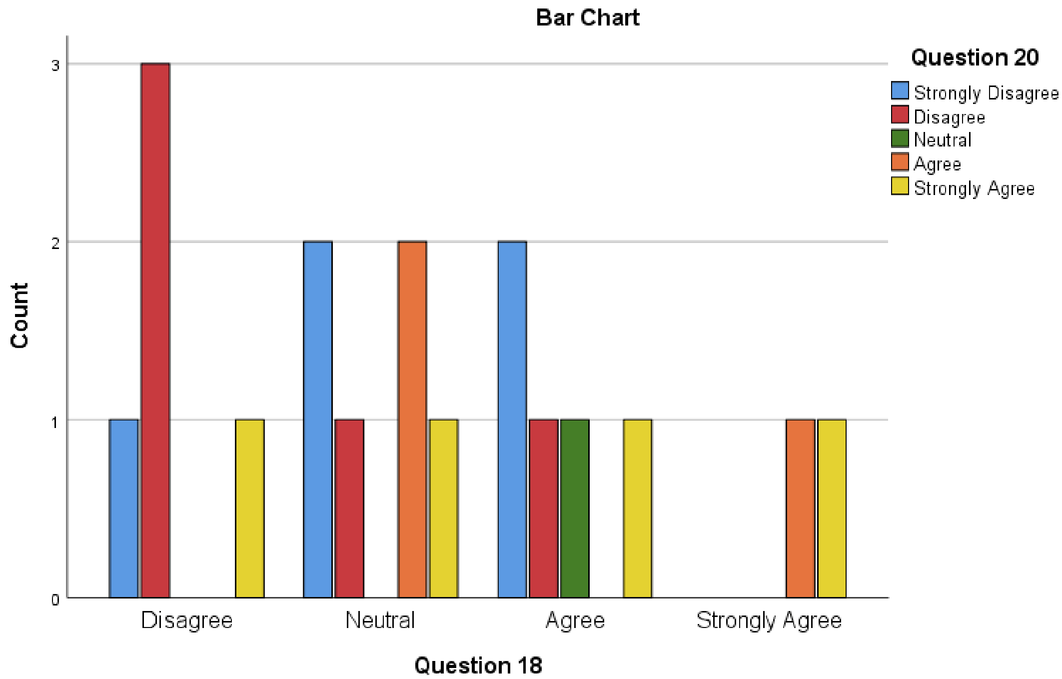

Figure 7.16: Distribution of system simplifications based on reusability of other objects

The results show that some users were able to benefit from existing simplifications in the system, but even more others were not, and an overwhelming number were not able to re-use other user's existing analytical objects. This users' inability to collaborate as expected in the system would need to be checked against their computer skills. A comparison is show below for questions 3 and 20:

Q3/Q20	1	2	3	4	5
1	0	3	0	0	2
2	0	0	1	0	0
3	1	1	0	0	1
4	2	0	0	1	0
5	2	1	0	2	1

Table 7.19: Distribution of computer skills based on system reusability

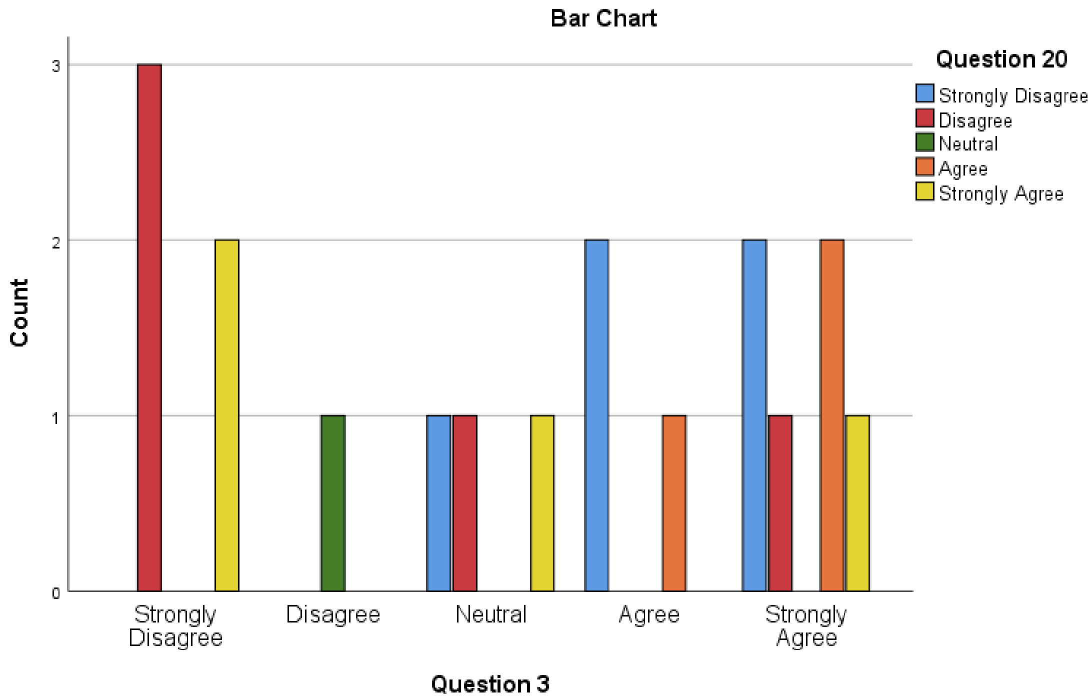

Figure 7.17: Distribution of computer skills based on system reusability

The results show that computer skills did not play an important role in helping users share objects among each other. This shows that the framework was not helping users much with sharing objects, although when used individually, resulted in acceptable results by the top management. Another comparison would be needed to show the dependency of the responses for question 21, where one's own queries were being reused. The table below shows this comparison:

Q3/Q21	1	2	3	4	5
1	2	1	1	1	0
2	0	0	1	0	0
3	0	0	1	1	1
4	0	0	0	1	2
5	2	0	0	2	2

Table 7.20: Comparing users' own reusability with computer skills

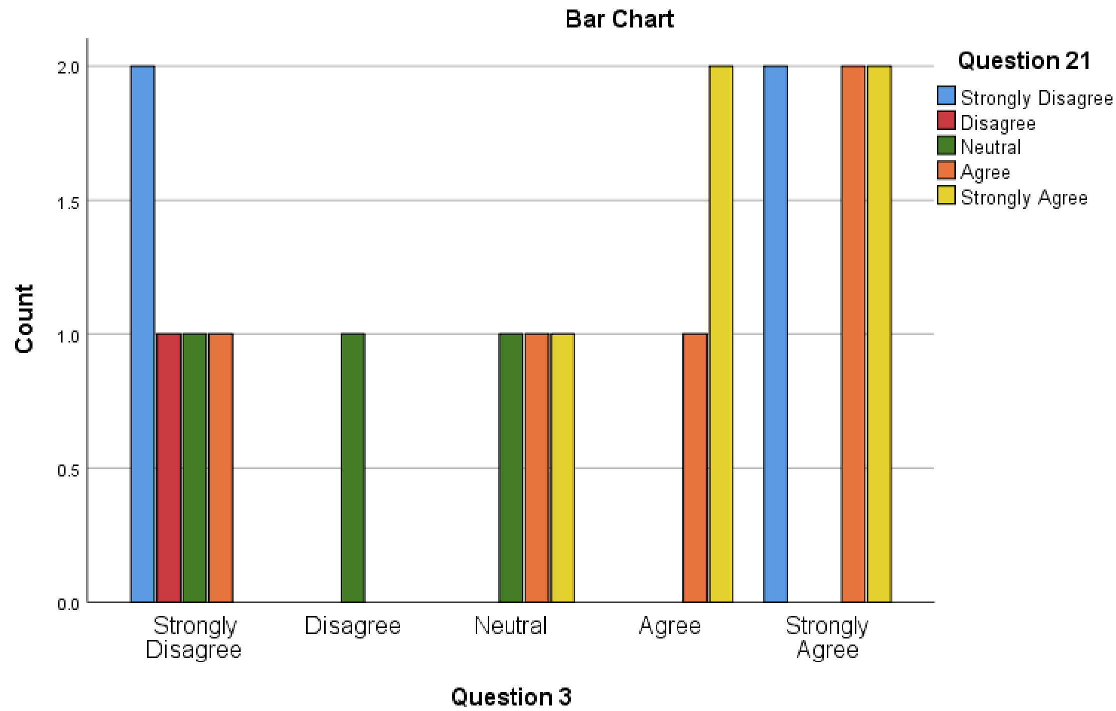

Figure 7.18: Comparing users' own reusability with computer skills

The results in this comparison are better than those in the previous one. This can be explained in such that the experienced users could individually reuse existing work that they did before, but that most users with moderate computer skills could not. Therefore, from a reusability perspective, the framework allows users to easily reuse their own functionalities but did not help much between different users. This should be taken further in future work as an improvement opportunity.

The last comparison would be the users' computer skills with their ability to share the own objects with other users, which is described in the statement from question 22:

Q3/Q22	1	2	3	4	5
1	0	0	1	3	1
2	0	0	0	1	0
3	0	1	0	1	1
4	1	0	2	0	0
5	0	4	0	1	1

Table 7.21: Computer skills to Object sharing ability Distribution

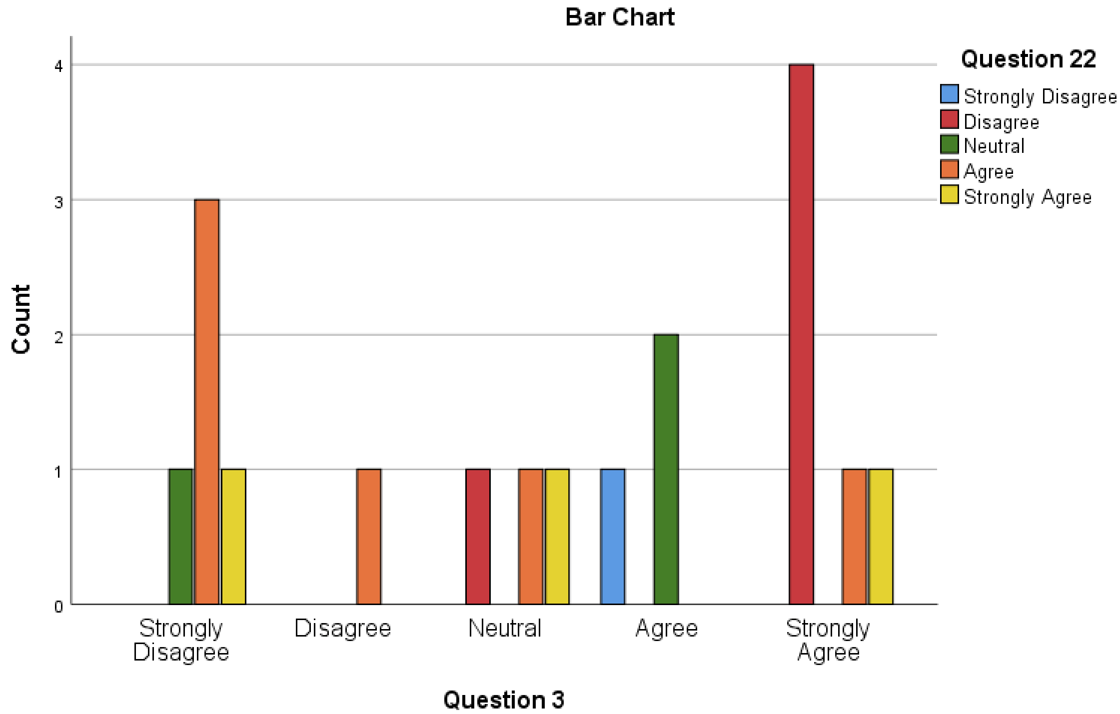

Figure 7.19: Computer skills to Object sharing ability Distribution

The table shows another potential weakness, where the majority of skilled users did not communicate their objects easily over the system with other users. This is also a point to be taken on for future enhancements, as users would probably reduce the effort spent doing this if the objects that other users produce could be easily communicated to them. Finally, a summary of each section's assessments is shown below. This will help focus the grouping of responses based on sections and not questions. It is assumed that the statements assessed by the users in the questionnaire are properly grouped in the corresponding sections, and the following graph gives an overall identification of the parts of the deliverable functionalities that would require further improvements:

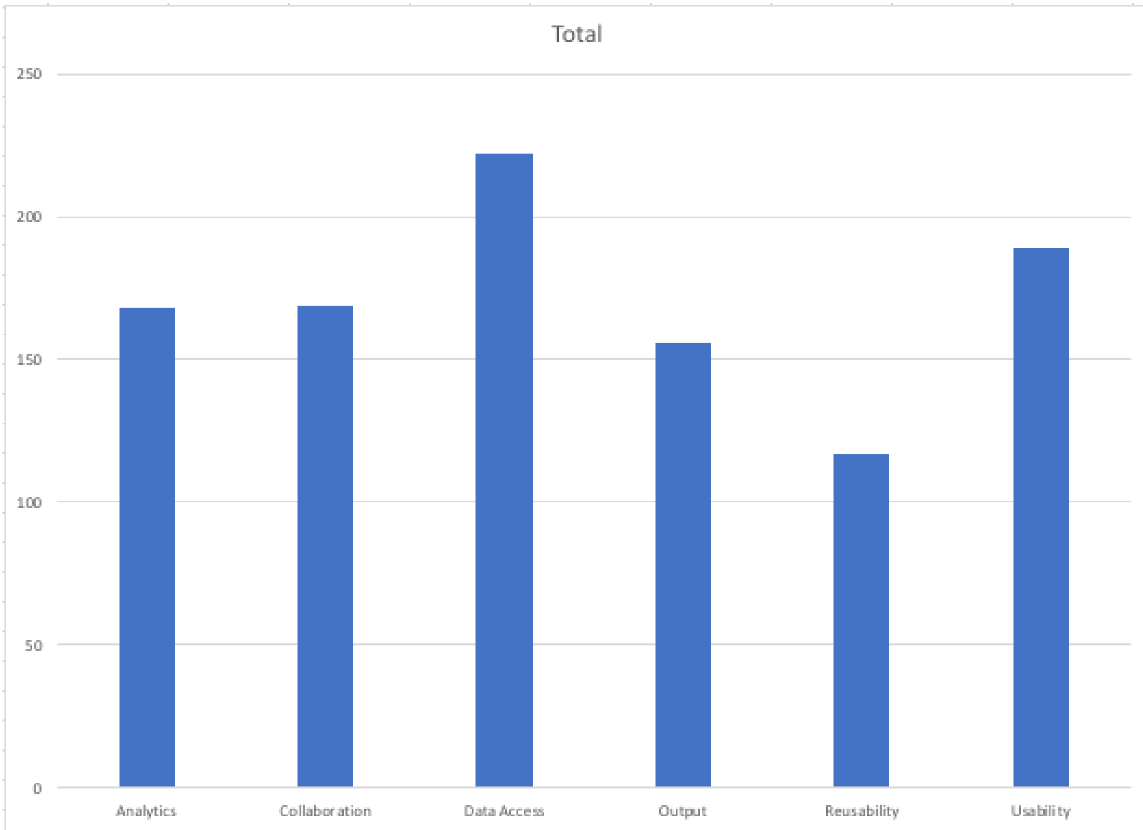

Figure 7.20: Total score of questionnaire sections

The figure confirms in a visual way the results that were discussed in the previous tabular distributions, where the statements in the reusability section received the lowest verification levels, showing that the framework system has a reusability gap that should be improved in future works. The responses showed that the most accurate statements of the questionnaire sections were related to the "Data Access" section, where users confirmed the ability of the system to retrieve and organize the information that is in the underlying healthcare setting systems. The system was also confirmed by the users to have an acceptable level of analysis and usability, both of which are main objectives in any software system implementation.

To complete the evaluation process, and in compliance with the methodology presented in Chapter 2 and in the literature review section of this chapter, a qualitative evaluation was conducted following the results presented in this section. The qualitative evaluation included follow-up meetings with user-groups that evaluated the system in a non-positive response, namely the middle management group with moderate to little computer skills. The group from different departments were asked to contribute verbally the ways that they feel would improve the system to be better accepted when it comes to performing analytical tasks. The users expressed their understanding of the importance of delivering summarized and aggregated results to the top management and the lack of time needed to manually do this. Among the most notable suggestions were ideas related to templated analytical routines, where a proven analytical approach is compiled into a template that the user can simply select and feed it the updated data.

The users who undertook the evaluation questionnaire had some interesting remarks that they expressed during and after the interviews and questionnaire submission. The users expressed that it is not very easy to find some objects in the system such as identities when they are loaded in large numbers into the system. The users mentioned examples such as Patients and Drugs, where many records might match a single string search. The users suggested that these records be made easier to find through an advanced search involving their and other's pre-defined analytics. An example for finding a patient is to enable the identities search list with user analytics such as "Search within Patients that had a surgery last year" or similar advanced searches.

Users also expressed their inability to navigate the framework's web interface easily, and suggested some ergonomic enhancemnts such as using a more modern web-interface and larger buttons that are visible on the form when the screen is small, such as a tablet. Since the framework was used also in clinics to understand patients' intersecting cases and diagnoses, the clinicians suggested to add more alerts and email notifications that enable a doctor to get some facts without opening the web application.

7.5 Summary, conclusion and discussions

This chapter started with a specific review of the software systems' evaluation methodologies and techniques, which resulted in a choice of a hybrid assessment approach where a quantitative questionnaire-based approach is mixed with a qualitative one-on-one approach where the users are addressed in person to explain their acceptance and rejection of the software system under evaluation. The chapter then followed with an explanation of the evaluation approach, starting with meetings with different user-groups in the organization. The meetings resulted in a development of the questionnaire that was distributed to the users following an implementation phase where the users experienced the system and used it to facilitate the creation of a monthly status report. The users' responses were then analysed using a comparative approach where the acceptance of the statements under question were discussed based with a tabular correlation method. The chapter then concluded with a meeting with the users who contributed to the lowest evaluation of the system. This would then form the basis of future improvements to the system.

Chapter 8

Conclusions, Discussion and Future Work

8.1 Introduction

This chapter summarises the work presented in this thesis and sheds light on the non-technical aspects that accompanied the work. This thesis started with setting an aim and stating the objectives that would be reached for the study to be successful. The study presented through a literature review several state-of-the-art methods for achieving a similar aim, and then used a comparative approach to identify a set of gaps in the existing approaches. The study followed that by creating a set of missing features and functionalities in the current solutions to the existing problem that would be needed in order to consider the problem as solved. This thesis then presented a solution that was then developed in further chapters and evaluated through an application scenario.

This chapter considers the challenges and obstacles that were faced throughout the phases of this thesis and revisits the aim and objectives, focusing on the issues faced when achieving the objectives. The research problem is also revisited, and the proposed solution is compared to the research problem initially stated. The chapter concludes with a summary of the issues related to the work designed and how to improve it in the future.

8.2 Discussion of clinical analytics' adoption in healthcare

Clinical analytics has helped transform healthcare data into viable knowledge that improves the well-being of patients and the overall quality of healthcare providers. The ability of an organization to adopt a clinical analytical approach as part of its patient-centric healthcare processes is critical to the progress of this organization's healthcare profile as more data, patients and challenges come along. The ethical, operational and financial circumstances and challenges of adopting clinical analytics in a healthcare setting has been discussed in the literature as part of the survey and examination of various clinical analytical technologies such as big data and business intelligence and how they integrate in modern settings. In [59] the adoption of big-data into clinical heterogeneous settings is examined, and the work focuses on how disease-related investigations using big-data can contribute to improving patient health standards. The outcomes of disease heterogeneity detection are mentioned as a major success factor for the adoption of big-data. [59] surveyed the literature from a big-data healthcare adoption perspective and focused on the technologies, benefits and application of the techniques and technologies in a clinical analytical setting. The work selected the literature that was published in the last 5 years with a clear focus on big-data with a modern approach, avoiding literature with classical data analytics approaches. Following a categorization of 87 matching literature works, [59] concludes that the IT infrastructure deficiencies, the huge costs of analytical tools adoption, data security, privacy and ownership challenges, in addition to quality and integrity issues related to data were seen as the most influential factors affecting clinical analytical adoption in healthcare. The results are in alignment with other literature reviewed in Chapter 3 and Chapter 4.

The adoption challenges and the ethical, medical and management integration complexities were examined in the work of [42], particularly through the examination of clinical prediction rules (CPR's) in a healthcare setting. The work starts with a literature review of how the CPR's adoption is growing in healthcare settings compares to the overall adoption of prediction rules. The review found that CPR's have a higher adoption rate in the healthcare settings and are therefore improving the contribution of clinical analytics to the treatment and consultation processes. [42] goes to explain how the resistance to CPR's varies from domain to another with the same healthcare setting. The work shows that CPR's have a better adoption and acceptability rate in the emergency and specialized medical departments than in the general medicine ones. The reason was shown to be the reluctance of general practitioners to alter or stop their consultation processes or slow them down by responding to CPR findings that might or might not reflect a positive result. The work noted that the value of CPR's was found to be in the detection of slight risk differences such as the risk of cancer in patients. The same overall concerns mentioned in [42] are consistent with those in [59].

Adoption and embedding of clinical analytics in healthcare settings, both in terms of technology or processes, faces multiple risks and challenges that need to be seriously addressed in line with the discussions from Chapter 4 to avoid implementation failures and to reduce the impact of introducing a decision-support tool at the wrong stage of a patient’s consultation or diagnosis stages. As shown from the literature reviewed in this and previous chapters, the purpose of using clinical analytics in healthcare should be to improve and enhance processes that involve consumption of large volumes of data to reach a clinical decision, rather than integrate clinical analytics at every possible stage in the operational and medical processes.

8.3 Summary of Chapters and Contribution

In this thesis an attempt was made to explain a problem and a gap in the existing decision-support and clinical intelligence processes in the health informatics domain. The thesis started in Chapter 1 with a general introduction of the problem domain, followed by a layout of methodologies in Chapter 2 that will be used in subsequent chapters to approach different subjects such as the literature review or the analysis of requirements. Chapter 3 conducted a detailed literature review of the latest state-of-the-art approaches to solutions of similar problems in the healthcare and other domains. The chapter covered several domains of technology and analytics that have contributed to the growth of the healthcare informatics field. The chapter concludes with an identification of limitations that prevent the existing approaches from solving the problem at hand. This is further detailed in Chapter 4, where the limitations are further analysed and compared to come to a definitive conclusion that there is a need for a new system to solve the described problem. The solution would have to enable non-technical clinicians and decision-support personnel to create aggregations and analytic enquiries against the various data sets that are consumed from different information systems and medical devices and systems in the healthcare domain.

Chapter 5 performs a series of analysis actions to understand and establish the required actions needed to reach the aim of this thesis. The chapter starts with a case study of the existing environment, using a questionnaire methodology to ask the users of different hospitals and of different educational and professional backgrounds how they perceive the current situation and what challenges they face when performing their decision-support and data analytical study of the healthcare information. The chapter analyses the results of this case study and uses the soft systems methodology to extract the real-life problems into the analytical abstract world, solve them and go back to the real-world to apply this solution to the identified problems. The chapter concludes with establishing the foundation for the requirements and needs of the new approach.

Chapter 6 designs and develops this new framework system using the software methodologies explained in Chapter 2. The components and modules of the framework system are explained in detail, in addition to the method of creating them and integrating them based on best practices. The chapter concludes with a sold implementation of the framework software system as part of an application scenario where the users in the local hospital are asked to replace their manual work related to creating the monthly status report by the usage of the framework system developed in this thesis.

In Chapter 7 the users who used the framework system to perform their activities are asked to evaluate this experience and are thus asked in an evaluation questionnaire a series of questions related to their different exposures of the framework system. The outcome of this evaluation is analysed and presented in this chapter.

8.4 Conclusions

This thesis presented a framework system that helped clinicians, middle management, and top management in a healthcare organization to design, organize, and exchange derived knowledge from the underlying heterogeneous healthcare systems. This chapter revisits previously set aims, objectives and problem statements that were stated in previous chapters, and then analyses the impact of the system on the people, organization and global healthcare status.

8.4.1 Fulfillment of the Thesis Aim

The aim of this thesis was initially stated in Chapter 1, with objectives related to achieving this aim. The aim is that of creating a system in the form of a modular framework which allows the non-technical specialists in a healthcare organization to process and eventually extract knowledge from an ongoing process of data feeding and loading. The proposed framework in this thesis was designed following an extensive study of the existing solutions that have been published in the literature, objectively comparing the solutions and how they experience limitations that would otherwise be resolved with the proposed system. The limitations were identified in Chapter 4 following the outcomes of the literature review that was conducted in Chapter 3. The limitations were analysed using methodologies that were presented in Chapter 2, and the system design was further developed, implemented and evaluated in Chapter 6 and Chapter 7. The objectives presented in Chapter 1 were carried out into conclusion, and for each objective the work faced some obstacles that were not of a technical nature, but rather related to social, economic and organizational reasons.

8.4.2 Objectives Fulfillment

The first objective to be achieved was: "Examine the current approaches in the healthcare informatics domain, identifying the most similar approaches and critically studying their existing solutions". This objective was achieved in the conclusion of Chapter 3, where the literature was reviewed based on a review methodology that was established in Chapter 2. The review provided a clear overview of the existing solutions that would be comparable to the one suggested by this thesis. The literature to be reviewed was mainly with online content that could be downloaded in full-text copies from online portals that offered affiliated login functions with Kingston University – London. It was sometimes the case that a full-text was found that would contribute to the value of the literature review but was not possible to download due to the lack of associated login credentials. For those cases, the full text was either purchased from the offering portal or requested in person from the main authors.

The second objective to be achieved was: “Identify the need for a new approach, based on the limitations of the existing approaches”. The limitations were identified in Chapter 4 based on a comparative review of the solutions previously stated. The objective was achieved with the identification of these limitations but faced a challenge of looking into every solution and analysing the published outcomes based on the method of solution presented in each work reviewed. Since no unified platform exists to compare healthcare informatics solutions with the same approach and design as the one presented in this thesis, the solutions were benchmarked based on globally accepted indicators that are often used in healthcare environments and based on personal experience in the healthcare setting.

The third objective to be achieved was: “Analyse the current approaches and the proposed approach”. The objective was achieved through the analysis performed in Chapter 5, where a combination of soft and hard analytical methodologies was applied to the outcomes of the previous chapters, namely Chapter 3 and Chapter 4. During the analysis a challenge was faced that the existing solutions and approaches were not all of a similar design complexity, namely the wide range of differences in implementation and technologies involved. While some solutions employed out-of-the-shelf technologies and tools that are easily accessible to researchers, others used proprietary or uncommon platforms and tools. The analysis of different approaches to the same problem is usually possible only through standard sets of proven measurable indicators. The lack of such a standard framework for objective analysis makes it difficult for researchers to correctly and accurately assign a rank for the different aspects of the solutions reviewed. Nonetheless, the approaches were analysed to the most accurate level based on domain knowledge, deliverable outputs, and its ability to extend and accommodate new technology.

The fourth objective was: “Design the new system and then and develop it, following a methodological approach and software development best-practices.”. The challenge faced was in the design and development of a solution that could consume data from different data sources in a healthcare setting. The data sources would not be known to the framework ahead of time, and it would be the task of the clinician to interact with the framework and define the data models that govern the incoming data. The challenge was to create a usable meta modelling user-interface that would be acceptable by non-technical users. Users expect medical terminologies and calculations to be available in the tool that would help in their analytical tasks, and those terms need to be fed into the framework to allow easier definition of the clinicians’ data models. The framework was built from ground-up with minimal pre-defined language-sets, and therefore required an intermediate setup that would familiarize the users with the tool. Although users could define their own groups and meta structures for their analytic, most users would retain from using the tool if they had to do that. In addition, it was found that users tend to duplicate existing definitions just because they did not create them themselves. Most users found it convenient to start using the tool with existing definitions, and to avoid any duplications of the definitions, a quality check needed to be done against the existing setup to make sure that terms that are relevant to the healthcare setting and in the most part already created.

8.4.3 Other Challenges

Software acceptance was another challenge that was faced in the adoption of such a tool in the healthcare environment. Most users come from a background where similar tools were used, either as part of a clinical system or standalone. The users needed to be exposed to the functionality of the new framework in such a way that would not lead them to abandoning the new experience for the sake of complexity. The challenges were of social nature, because the users were almost always used to getting assistance in the form of technical support from someone else. In the adoption of such a framework the users were using the system directly, instructing the framework to consume existing data and applying pre-defined meta models to it. Another challenge was the difficulty to convince clinicians that once their analysis was validated against the traditional way of delivering knowledge to their top management, the next month and later on the same output would be automatically valid, and they would not need further confirmations from any technical person.

Top management and stakeholders were also involved in this social change. Their usual behaviour was that once they receive a report from a non-technical middle management post, they would then address it to the technical department to make sure that the numbers are correct. The technical department would – most of the time – go back to the middle management personnel and ask for clarifications on the intended analysis and would re-run the data computations to retrieve similar results.

8.4.4 Impacts of the Framework System

Although the framework presented in this thesis faced the aforementioned challenges before, during and after its design and development, the impact that the framework system created in the environment that it was delivered to was significant. The framework contributed to reducing the analytical overhead required to generate a monthly report for the top management. The framework was not easy to adopt by users and needed further assistance during the introduction phase, but then was able to serve as a main delivery platform for further analytical requirements. The platform helped the organization gain further control of its data by creating a workbench where top and middle management from both medical and administrative departments could work together and create their experiments and improve them over time to reach a self-created business and analytical output. The solution which combines both the framework and the people who use it to serve knowledge from healthcare data defined a new approach to how healthcare personnel deliver outcome to their customers. In recent models, the clinicians who were involved in analysis of healthcare data had to spend time to redo normal repetitive tasks of ordering data, collecting it, formatting it, loading it, and then performing the built-in hard-coded conversions that put the data in a format that could be presented to other users. This time was taken away from other more important activities that are part of the core specialty of the healthcare personnel. An example is found at a nursing station, where a monthly nursing activity report is requested by the quality department to assess nurse-to-patient ratios. While the healthcare information system provides patient status and occupation rates, the combination computation is not easy, and would need to be calculated for every patient.

The implementation of the framework system did not have a direct impact on patient safety, but it did help in improving the quality of knowledge delivery to the top management, which improved the ability of the administrative and middle management to coordinate better and delivery higher quality of service to their patients, which indirectly resulted in improved patient safety.

8.5 Discussion

The work done in this thesis included several non-technical topics that needed to be addressed and was distributed across several phases of different requirements and challenges. The first part of the work started with a review of existing literature where the topic of decision support in medical informatics was researched. The main challenge was the lack of local research due to the fact that the topic of medical informatics in general was new at the time when the study started. The research was therefore automatically redirected toward large healthcare organizations and groups that have interest in medical informatics and apply its discipline in daily activities and processes.

The research expanded to include other topics of interest to the main decision-support in medical informatics. Most of the work encountered lead to other works of similar but indirect relationship to the original source. As the research expanded, the interest was then renewed in decision-support in medical informatics as the local healthcare organization expanded and started requesting analytical reports and basing decisions in the organization on the outcomes of such reports. The need expanded further as units and departments started gaining interest in local analytics to help them explain the overall status of their performance as a unit and as individuals. Later into the study, doctors and clinics within the healthcare organization started to be interested in data informatics to improve their efficiency with patient case handling. The overall need for analytical decision-support was officially established through the creation of an independent department that had the responsibility of delivering knowledge and not technology.

In the process of preparing the work for this thesis, the framework system was developed in phases that contributed to the final product. The initial design decision was to build the framework on-top of a solid web API model that makes it easy for future integrations with other systems. The second design decision was to make it a service, not an application, which required that it be runnable on operating systems that are easily deployed to cloud appliances. The choice of the programming language was based on previous experience.

8.6 Future Work

The next steps in the research that was started in this thesis is to enhance the framework that was developed and to make it capable of operating in larger and more complex data ecosystems, where the challenges are distributed across multiple domains and divisions and where the data consumers or the stakeholders have vaster interest scopes and require advanced predictability based on analytical methods. The main topics of enhancement would be improving the user-interface and in therefore the user experience through adoption of modern user-interface libraries that could present results and interact with the user in a smoother and more efficient way. The other aspects in the framework would also need improvement based on the feedback from the users in the evaluation survey performed in Chapter 7, namely as related to sharing results between users and making the data-loading process faster through improving the existing import algorithms.

A continuous evaluation methodology would also need to be adopted so that the framework system would maintain feedback links from the users where it is being implemented, in a consistent manner and in application of the continuous evaluation practices of software systems from the evaluation literature review in Chapter 2. In conclusion, the adoption and utilization of a software system in general and a framework system in particular is a continuous process where the problems faced during the user adoption stages are later resolved and fed-back into the main development and design processes to ensure a continuously improving software system in both its current and future domains of operation.

References

1. Ahmadian, L., Salehi Nejad, S. and Khajouei, R. (2015). Evaluation methods used on health information systems (HISs) in Iran and the effects of HISs on Iranian healthcare: A systematic review. International Journal of Medical Informatics, 84(6), pp.444-453.

2. Antoniadis, I., Tsiakiris, T. and Tsopogloy, S. (2015). Business Intelligence During Times of Crisis: Adoption and Usage of ERP Systems by SMEs. Procedia - Social and Behavioral Sciences, 175, pp.299-307.

3. Baird, R. (2018). Systematic reviews and meta-analytic techniques. Seminars in Pediatric Surgery, 27(6), pp.338-344.

4. Barki, H. and Hartwick, J. (1994). Measuring User Participation, User Involvement, and User Attitude. MIS Quarterly, 18(1), p.59.

5. Bates, D., Kuperman, G., Wang, S., Gandhi, T., Kittler, A., Volk, L., Spurr, C., Khorasani, R., Tanasijevic, M. and Middleton, B. (2003). Ten Commandments for Effective Clinical Decision Support: Making the Practice of Evidence-based Medicine a Reality. Journal of the American Medical Informatics Association, 10(6), pp.523-530.

6. Beck, B. (2012). BioProfile—Extract knowledge from corporate databases to assess cross-reactivities of compounds. Bioorganic & Medicinal Chemistry, 20(18), pp.5428-5435.

7. Belgrave, D., Henderson, J., Simpson, A., Buchan, I., Bishop, C. and Custovic, A. (2017). Disaggregating asthma: Big investigation versus big data. Journal of Allergy and Clinical Immunology, 139(2), pp.400-407.

8. Booch, G., Rumbaugh, J. and Jacobson, I. (2005). The unified modeling language user guide. Upper Saddle River, NJ: Addison-Wesley.

9. Bose, R. (2003). Knowledge management-enabled health care management systems: capabilities, infrastructure, and decision-support. Expert Systems with Applications, 24(1), pp.59-71.

10. Boynton, P. and Greenhalgh, T. (2004). Selecting, designing, and developing your questionnaire. BMJ, 328(7451), pp.1312-1315.

11. Brichni, M., Dupuy-Chessa, S., Gzara, L., Mandran, N. and Jeannet, C. (2017). BI4BI: A continuous evaluation system for Business Intelligence systems. Expert Systems with Applications, 76, pp.97-112.

12. Campbell, M. (2000). Framework for design and evaluation of complex interventions to improve health. BMJ, 321(7262), pp.694-696.

13. Chae, Y., Kim, H., Tark, K., Park, H. and Ho, S. (2003). Analysis of healthcare quality indicator using data mining and decision support system. Expert Systems with Applications, 24(2), pp.167-172.

14. Cherednichenko, O. and Yangolenko, O. (2013). Towards Quality Monitoring and Evaluation Methodology: Higher Education Case-Study. Lecture Notes in Business Information Processing, pp.120-127.

15. Chiaramello, E., Pinciroli, F., Bonalumi, A., Caroli, A. and Tognola, G. (2016). Use of “off-the-shelf” information extraction algorithms in clinical informatics: A feasibility study of MetaMap annotation of Italian medical notes. Journal of Biomedical Informatics, 63, pp.22-32.

16. Conde-Clemente, P., Trivino, G. and Alonso, J. (2017). Generating automatic linguistic descriptions with big data. Information Sciences, 380, pp.12-30.

17. Danciu, I., Cowan, J., Basford, M., Wang, X., Saip, A., Osgood, S., Shirey-Rice, J., Kirby, J. and Harris, P. (2014). Secondary use of clinical data: The Vanderbilt approach. Journal of Biomedical Informatics, 52, pp.28-35.

18. Davies, J., Gibbons, J., Harris, S. and Crichton, C. (2014). The CancerGrid experience: Metadata-based model-driven engineering for clinical trials. Science of Computer Programming, 89, pp.126-143.

19. Dimeglio, C., Kelly-Irving, M., Lang, T. and Delpierre, C. (2016). Expectations and boundaries for Big Data approaches in social medicine. Journal of Forensic and Legal Medicine.

20. Donald Wynn and Clay K. Williams. Principles for conducting critical realist case study research in information systems. MIS Q. 36, 3 (September 2012), 787-810.

21. Duffy, B., Zhu, L., Decornez, H. and Kitchen, D. (2012). Early phase drug discovery: Cheminformatics and computational techniques in identifying lead series. Bioorganic & Medicinal Chemistry, 20(18), pp.5324-5342.

22. El Naqa, I. (2016). Perspectives on making big data analytics work for oncology. Methods, 111, pp.32-44.

23. Elwyn, G. (2006). Developing a quality criteria framework for patient decision aids: online international Delphi consensus process. BMJ, 333(7565), pp.417-0.

24. Elwyn, G., Scholl, I., Tietbohl, C., Mann, M., Edwards, A., Clay, C., Légaré, F., Weijden, T., Lewis, C., Wexler, R. and Frosch, D. (2013). “Many miles to go …”: a systematic review of the implementation of patient decision support interventions into

routine clinical practice. BMC Medical Informatics and Decision Making, 13(Suppl 2), p.S14.

25. Eslami Andargoli, A., Scheepers, H., Rajendran, D. and Sohal, A. (2017). Health information systems evaluation frameworks: A systematic review. International Journal of Medical Informatics, 97, pp.195-209.

26. Farhood, B., Geraily, G. and Abtahi, S. (2019). A systematic review of clinical applications of polymer gel dosimeters in radiotherapy. Applied Radiation and Isotopes, 143, pp.47-59.

27. Forkan, A., Khalil, I. and Atiquzzaman, M. (2017). ViSiBiD: A learning model for early discovery and real-time prediction of severe clinical events using vital signs as big data. Computer Networks, 113, pp.244-257.

28. Gaardboe, R., Nyvang, T. and Sandalgaard, N. (2017). Business Intelligence Success applied to Healthcare Information Systems. Procedia Computer Science, 121, pp.483-490.

29. Gastaldi, L., Pietrosi, A., Lessanibahri, S., Paparella, M., Scaccianoce, A., Provenzale, G., Corso, M. and Gridelli, B. (2018). Measuring the maturity of business intelligence in healthcare: Supporting the development of a roadmap toward precision medicine within ISMETT hospital. Technological Forecasting and Social Change, 128, pp.84-103.

30. Grover, V. and Lyytinen, K. (2015). New State of Play in Information Systems Research: The Push to the Edges. MIS Quarterly, 39(2), pp.271-296.

31. Gu, D., Li, J., Li, X. and Liang, C. (2017). Visualizing the knowledge structure and evolution of big data research in healthcare informatics. International Journal of Medical Informatics, 98, pp.22-32.

32. Hanson, J., Melnyk, S. and Calantone, R. (2011). Defining and measuring alignment in performance management. International Journal of Operations & Production Management, 31(10), pp.1089-1114.

33. Hettinger, A., Roth, E. and Bisantz, A. (2017). Cognitive engineering and health informatics: Applications and intersections. Journal of Biomedical Informatics, 67, pp.21-33.

34. Hu, H., Liang, W., Liu, M., Li, L., Li, Z., Li, T., Wang, J., Shi, T., Han, S., Su, M., Peng, X., Peng, Y., Zhao, W., Wang, B., Zhang, P. and Zhu, W. (2010). Establishment and evaluation of a model of a community health service in an underdeveloped area of China. Public Health, 124(4), pp.206-217.

35. Hui, J., Knoop, S. and Schwarz, P. (2012). Analyzing clinical data in XML. Proceedings of the 2nd ACM SIGHIT symposium on International health informatics - IHI '12.

36. Johnston, A., Kelly, S., Hsieh, S., Skidmore, B. and Wells, G. (2019). Systematic reviews of clinical practice guidelines: a methodological guide. Journal of Clinical Epidemiology, 108, pp.64-76.

37. Kaipio, J., Lääveri, T., Hyppönen, H., Vainiomäki, S., Reponen, J., Kushniruk, A., Borycki, E. and Vänskä, J. (2017). Usability problems do not heal by themselves: National survey on physicians' experiences with EHRs in Finland. International Journal of Medical Informatics, 97, pp.266-281.

38. Kambatla, K., Kollias, G., Kumar, V. and Grama, A. (2014). Trends in big data analytics. Journal of Parallel and Distributed Computing, 74(7), pp.2561-2573.

39. Kannampallil, T., Abraham, J. and Patel, V. (2016). Methodological framework for evaluating clinical processes: A cognitive informatics perspective. Journal of Biomedical Informatics, 64, pp.342-351.

40. Kao, H., Yu, M., Masud, M., Wu, W., Chen, L. and Wu, Y. (2016). Design and evaluation of hospital-based business intelligence system (HBIS): A foundation for design science research methodology. Computers in Human Behavior, 62, pp.495-505.

41. Kawamoto, K., Houlihan, C., Balas, E. and Lobach, D. (2005). Improving clinical practice using clinical decision support systems: a systematic review of trials to identify features critical to success. BMJ, 330(7494), p.765.

42. Kennedy, G. and Gallego, B. (2019). Clinical prediction rules: A systematic review of healthcare provider opinions and preferences. International Journal of Medical Informatics, 123, pp.1-10.

43. Khajouei, R., Zahiri Esfahani, M. and Jahani, Y. (2016). Comparison of heuristic and cognitive walkthrough usability evaluation methods for evaluating health information systems. Journal of the American Medical Informatics Association, p.ocw100.

44. Khalifa, M. (2015). Reducing Emergency Department Crowding Using Health Analytics Methods: Designing AnEvidence Based Decision Algorithm. Procedia Computer Science, 63, pp.409-416.

45. Kong, L., Liu, Y., Li, G., Fang, Y., Kang, X. and Li, P. (2016). Resilience moderates the relationship between emotional intelligence and clinical communication ability

among Chinese practice nursing students: A structural equation model analysis. Nurse Education Today, 46, pp.64-68.

46. Kranjc, J., Orač, R., Podpečan, V., Lavrač, N. and Robnik-Šikonja, M. (2017). ClowdFlows: Online workflows for distributed big data mining. Future Generation Computer Systems, 68, pp.38-58.

47. Kriglstein, S., Leitner, M., Kabicher-Fuchs, S. and Rinderle-Ma, S. (2016). Evaluation Methods in Process-Aware Information Systems Research with a Perspective on Human Orientation. Business & Information Systems Engineering, 58(6), pp.397-414.

48. Kuperman, G., Bobb, A., Payne, T., Avery, A., Gandhi, T., Burns, G., Classen, D. and Bates, D. (2007). Medication-related Clinical Decision Support in Computerized Provider Order Entry Systems: A Review. Journal of the American Medical Informatics Association, 14(1), pp.29-40.

49. Lee, S., Kang, Y., Ialongo, N. and Prabhu, V. (2016). Predictive analytics for delivering prevention services. Expert Systems with Applications, 55, pp.469-479.

50. Légaré, F., O'Connor, A., Graham, I., Wells, G. and Tremblay, S. (2006). Impact of the Ottawa Decision Support Framework on the Agreement and the Difference between Patients' and Physicians' Decisional Conflict. Medical Decision Making, 26(4), pp.373-390.

51. Légaré, F., O’Connor, A. C., Graham, I., Saucier, D., Côté, L., Cauchon, M., & Paré, L. (2006). Supporting patients facing difficult health care decisions: Use of the Ottawa Decision Support Framework. Canadian Family Physician, 52(4), 476–477.

52. Légaré, F., O’Connor, A., Graham, I., Saucier, D., Côté, L., Blais, J., Cauchon, M. and Paré, L. (2006). Primary health care professionals’ views on barriers and facilitators

to the implementation of the Ottawa Decision Support Framework in practice. Patient Education and Counseling, 63(3), pp.380-390.

53. Li, Y., Jiang, W., Yang, L. and Wu, T. (2018). On neural networks and learning systems for business computing. Neurocomputing, 275, pp.1150-1159.

54. Liew, C. (2018). The future of radiology augmented with Artificial Intelligence: A strategy for success. European Journal of Radiology, 102, pp.152-156.

55. Lin, W., Dou, W., Zhou, Z. and Liu, C. (2015). A cloud-based framework for Home-diagnosis service over big medical data. Journal of Systems and Software, 102, pp.192-206.

56. Lu, Y., Xiao, Y., Sears, A. and Jacko, J. (2005). A review and a framework of handheld computer adoption in healthcare. International Journal of Medical Informatics, 74(5), pp.409-422.

57. Lukić, J., Radenković, M., Despotović-Zrakić, M., Labus, A. and Bogdanović, Z. (2016). A hybrid approach to building a multi-dimensional business intelligence system for electricity grid operators. Utilities Policy, 41, pp.95-106.

58. Marron, J. (2017). Big Data in context and robustness against heterogeneity. Econometrics and Statistics, 2, pp.73-80.

59. Mehta, N. and Pandit, A. (2018). Concurrence of big data analytics and healthcare: A systematic review. International Journal of Medical Informatics, 114, pp.57-65.

60. Moutinho, L. (2004). Soft Systems Methodology - Conceptual Model Building and Its Contribution. European Management Journal, 22(1), pp.123-124.

61. Mujtaba, G., Shuib, L., Idris, N., Hoo, W., Raj, R., Khowaja, K., Shaikh, K. and Nweke, H. (2019). Clinical text classification research trends: Systematic literature review and open issues. Expert Systems with Applications, 116, pp.494-520.

62. Parker, J., Summerfeldt, L., Taylor, R., Kloosterman, P. and Keefer, K. (2013). Problem gambling, gaming and Internet use in adolescents: Relationships with emotional intelligence in clinical and special needs samples. Personality and Individual Differences, 55(3), pp.288-293.

63. Patel, V. and Kaelber, D. (2014). Using aggregated, de-identified electronic health record data for multivariate pharmacosurveillance: A case study of azathioprine. Journal of Biomedical Informatics, 52, pp.36-42.

64. Perer, A. and Gotz, D. (2013). Data-driven exploration of care plans for patients. CHI '13 Extended Abstracts on Human Factors in Computing Systems on - CHI EA '13.

65. Perrons, R. and McAuley, D. (2015). The case for “n«all”: Why the Big Data revolution will probably happen differently in the mining sector. Resources Policy, 46, pp.234-238.

66. Peters, M., Wieder, B., Sutton, S. and Wakefield, J. (2016). Business intelligence systems use in performance measurement capabilities: Implications for enhanced competitive advantage. International Journal of Accounting Information Systems, 21, pp.1-17.

67. Petit, J., Meurice, N., Kaiser, C. and Maggiora, G. (2012). Softening the Rule of Five—where to draw the line?. Bioorganic & Medicinal Chemistry, 20(18), pp.5343-5351.

68. Phd, G., 2014. Clinical Intelligence: The Big Data Analytics Revolution In Healthcare: A Framework For Clinical And Business Intelligence. Createspace Independent Publishi.

69. Polyvyanyy, A., Ouyang, C., Barros, A. and van der Aalst, W. (2017). Process querying: Enabling business intelligence through query-based process analytics. Decision Support Systems, 100, pp.41-56.

70. Popovič, A., Hackney, R., Coelho, P. and Jaklič, J. (2014). How information-sharing values influence the use of information systems: An investigation in the business intelligence systems context. The Journal of Strategic Information Systems, 23(4), pp.270-283.

71. Prat, N., Comyn-Wattiau, I. and Akoka, J. (2015). A Taxonomy of Evaluation Methods for Information Systems Artifacts. Journal of Management Information Systems, 32(3), pp.229-267.

72. Presser, S., Couper, M., Lessler, J., Martin, E., Martin, J., Rothgeb, J. and Singer, E. (2004). Methods for Testing and Evaluating Survey Questions. Public Opinion Quarterly, 68(1), pp.109-130.

73. Radenković, M., Lukić, J., Despotović-Zrakić, M., Labus, A. and Bogdanović, Z. (2018). Harnessing business intelligence in smart grids: A case of the electricity market. Computers in Industry, 96, pp.40-53.

74. Raghupathi, W. and Raghupathi, V. (2014). Big data analytics in healthcare: promise and potential. Health Information Science and Systems, 2(1).

75. Rehman, M., Chang, V., Batool, A. and Wah, T. (2016). Big data reduction framework for value creation in sustainable enterprises. International Journal of Information Management, 36(6), pp.917-928.

76. Riazanov, A., Rose, G., Klein, A., Forster, A., Baker, C., Shaban-Nejad, A. and Buckeridge, D. (2012). Towards clinical intelligence with SADI semantic web services. Proceedings of the 4th International Workshop on Semantic Web Applications and Tools for the Life Sciences - SWAT4LS '11.

77. Richesson, R. (2014). An informatics framework for the standardized collection and analysis of medication data in networked research. Journal of Biomedical Informatics, 52, pp.4-10.

78. Rikhardsson, P. and Yigitbasioglu, O. (2018). Business intelligence & analytics in management accounting research: Status and future focus. International Journal of Accounting Information Systems, 29, pp.37-58.

79. Rodger, J. (2015). Discovery of medical Big Data analytics: Improving the prediction of traumatic brain injury survival rates by data mining Patient Informatics Processing Software Hybrid Hadoop Hive. Informatics in Medicine Unlocked, 1, pp.17-26.

80. Schumacher, A., Rujan, T. and Hoefkens, J. (2014). A collaborative approach to develop a multi-omics data analytics platform for translational research. Applied & Translational Genomics, 3(4), pp.105-108.

81. Shen, C., Chang, R., Hsu, C. and Chang, I. (2017). How business intelligence maturity enabling hospital agility. Telematics and Informatics, 34(1), pp.450-456.

82. Shiraishi, J., Li, Q., Appelbaum, D. and Doi, K. (2011). Computer-Aided Diagnosis and Artificial Intelligence in Clinical Imaging. Seminars in Nuclear Medicine, 41(6), pp.449-462.

83. Sim, I., Gorman, P., Greenes, R. A., Haynes, R. B., Kaplan, B., Lehmann, H., & Tang, P. C. (2001). Clinical Decision Support Systems for the Practice of Evidence-based Medicine. Journal of the American Medical Informatics Association : JAMIA, 8(6), 527–534.

84. Sligo, J., Gauld, R., Roberts, V. and Villa, L. (2017). A literature review for large-scale health information system project planning, implementation and evaluation. International Journal of Medical Informatics, 97, pp.86-97.

85. Sonntag, D., Tresp, V., Zillner, S., Cavallaro, A., Hammon, M., Reis, A., Fasching, P., Sedlmayr, M., Ganslandt, T., Prokosch, H., Budde, K., Schmidt, D., Hinrichs, C., Wittenberg, T., Daumke, P. and Oppelt, P. (2015). The Clinical Data Intelligence Project. Informatik-Spektrum, 39(4), pp.290-300.

86. Sookhak, M., Gani, A., Khan, M. and Buyya, R. (2017). Dynamic remote data auditing for securing big data storage in cloud computing. Information Sciences, 380, pp.101-116.

87. Teixeira, L., Santos, B., Saavedra, V. and Ferreira, C. (2017). Evaluation Methods to Support Health Information Systems Development: A Framework Supported in Literature and Practical Experience. Digital Human Modeling. Applications in Health, Safety, Ergonomics, and Risk Management: Health and Safety, pp.79-90.

88. Uçaktürk, A., Uçaktürk, T. and Yavuz, H. (2015). Possibilities of Usage of Strategic Business Intelligence Systems Based on Databases in Agile Manufacturing. Procedia - Social and Behavioral Sciences, 207, pp.234-241.

89. Vanderwall, D., Yuen, N., Al-Ansari, M., Bailey, J., Fram, D., Green, D., Pickett, S., Vitulli, G., Luengo, J. and Almenoff, J. (2011). Molecular clinical safety intelligence: a system for bridging clinically focused safety knowledge to early-stage drug discovery – the GSK experience. Drug Discovery Today, 16(15-16), pp.646-653.

90. Vennila, V. and Rajiv Kannan, A. (2016). Symmetric Matrix-based Predictive Classifier for Big Data computation and information sharing in Cloud. Computers & Electrical Engineering, 56, pp.831-841.

91. Vogt, M. and Bajorath, J. (2012). Chemoinformatics: A view of the field and current trends in method development. Bioorganic & Medicinal Chemistry, 20(18), pp.5317-5323.

92. Wang, Y., Kung, L. and Byrd, T. (2018). Big data analytics: Understanding its capabilities and potential benefits for healthcare organizations. Technological Forecasting and Social Change, 126, pp.3-13.

93. Ward, C., Wilson, J. and Fischer, R. (2011). Assessing the predictive validity of cultural intelligence over time. Personality and Individual Differences, 51(2), pp.138-142.

94. Ward, M., Marsolo, K. and Froehle, C. (2014). Applications of business analytics in healthcare. Business Horizons, 57(5), pp.571-582.

95. Warnars, H. (2011). Object-oriented modelling with unified modelling language 2.0 for simple software application based on agile methodology. Behaviour & Information Technology, 30(3), pp.293-307.

96. Wason, R. (2018). Deep learning: Evolution and expansion. Cognitive Systems Research, 52, pp.701-708.

97. Weiss, S., Indurkhya, N. and Apte, C. (2010). Predictive rule discovery from electronic health records. Proceedings of the ACM international conference on Health informatics - IHI '10.

98. Westra, B., Sylvia, M., Weinfurter, E., Pruinelli, L., Park, J., Dodd, D., Keenan, G., Senk, P., Richesson, R., Baukner, V., Cruz, C., Gao, G., Whittenburg, L. and Delaney, C. (2017). Big data science: A literature review of nursing research exemplars. Nursing Outlook, 65(5), pp.549-561.

99. Wheeler, F. and Checkland, P. (2000). Systems Thinking, Systems Practice: Includes a 30-Year Retrospective. The Journal of the Operational Research Society, 51(5), p.647.

100. Wols, A., Scholte, R. and Qualter, P. (2015). Prospective associations between loneliness and emotional intelligence. Journal of Adolescence, 39, pp.40-48.

101. Xie, T. (2017). Heteroscedasticity-robust model screening: A useful toolkit for model averaging in big data analytics. Economics Letters, 151, pp.119-122.

102. Yacci, M. (1997). Object-oriented compiler construction [Book Review]. IEEE Software, 14(5), pp.117-118.

103. Yang, P., Pan, F., Liu, D. and Xu, Y. (2012). Function Model for Community Health Service Information. Physics Procedia, 33, pp.167-171.

104. Yuan, Y., Xun, G., Suo, Q., Jia, K. and Zhang, A. (2019). Wave2Vec: Deep representation learning for clinical temporal data. Neurocomputing, 324, pp.31-42.

105. Zhao, R., Yan, R., Chen, Z., Mao, K., Wang, P. and Gao, R. (2019). Deep learning and its applications to machine health monitoring. Mechanical Systems and Signal Processing, 115, pp.213-237.

Appendix A

The following extract from such a system-defined method shows how content-types are handled in the returned structure.

```
<?
if(responseType==null || responseType=="" || responseType=="js")
{
?>
var ret = new Array();
<?=DataMap.toJS((DataMap[])((object[])oret)[1])?>

ret.Count = <?=(((object[])oret).Length==0?0:((object[])oret)[0])?>;<?
}
else if(responseType=="xml")
{
?><?=DataMap.ToXML((DataMap[])((object[])oret)[1])?><?
}
else if(responseType=="json")
{
?><?=DataMap.toJSON((DataMap[])((object[])oret)[1])?><?
}
?>
```

Figure 0.1: Example content-type Return

In the example above, the C# object array “oret” is converted to XML, JSON and a POJO content based on the “responseType” variable that is provided by the executing context. When the response-type is XML, the endpoint returns the following output example:

```xml
<?xml version="1.0" encoding="UTF-8"?>
<DataMapList>
    <DataMap>
        <Active type="bool"><![CDATA[True]]></Active>
        <Code type="string"><![CDATA[LABEX]]></Code>
        <Date type="DateTime"><![CDATA[Oct/04/2015 06:02:34]]></Date>
        <DateFormat type="string"><![CDATA[M/d/yyyy h:mm:ss tt]]></DateFormat>
        <DateKey type="string"><![CDATA[Trh_DateTime]]></DateKey>
        <Enabled type="bool"><![CDATA[True]]></Enabled>
        <Name type="string"><![CDATA[Lab Map]]></Name>
        <Remark type="string" />
        <TimeFormat type="string" />
        <TimeKey type="string" />
        <Id type="long"><![CDATA[1]]></Id>
        <Organization class="Organization">
            <Active type="bool"><![CDATA[True]]></Active>
            <Code type="string"><![CDATA[ARZ]]></Code>
            <Date type="DateTime"><![CDATA[Oct/04/2015 05:15:49]]></Date>
            <Enabled type="bool"><![CDATA[True]]></Enabled>
            <Name type="string"><![CDATA[[redacted]]]></Name>
            <Order type="int"><![CDATA[1]]></Order>
            <Id type="long"><![CDATA[1]]></Id>
            <DataSources array="true" />
            <Measures array="true" />
            <DataMaps array="true" />
            <Aggregations array="true" />
            <Groups array="true" />
        </Organization>
        <DataSets array="true" />
        <DataMapValues array="true" />
        <DataFields array="true" />
        <Batches array="true" />
    </DataMap>
    <DataMap>
        <Active type="bool"><![CDATA[True]]></Active>
        <Code type="string"><![CDATA[PTSALE]]></Code>
```


Figure 0.2: API Response as XML

```
[
  {
    "ToString":"Lab Map",
    "__LOADED":true,
    "__ROWID":1147342714,
    "Active":true,
    "Code":"LABEX",
    "Date":"\/Date(635795785540000000)\/",
    "DateFormat":"M/d/yyyy h:mm:ss tt",
    "DateKey":"Trh_DateTime",
    "Enabled":true,
    "Name":"Lab Map",
    "Remark":"",
    "TimeFormat":"",
    "TimeKey":"",
    "Id":1,
    "Organizationid":1,
    "Organization":{
      "ToString":"[redacted],
      "__LOADED":true,
      "__ROWID":745948491,
      "Active":true,
      "Code":"ARZ",
      "Date":"\/Date(635795757490000000)\/",
      "Enabled":true,
```


Figure 0.3: API Response as JSON

```
var method_name = "CorporateMeasures.comDataMapFindall";

var server_time = new Date(2017, 9 - 1, 12, 12, 24, 42);

var execution_time = 0.059652;


var ret = new Array();
ret = new Array();
ret[0] = new Object();
ret[0].ToString = "Lab Map";
ret[0]._ToString = "Lab Map";
ret[0].__LOADED = new Date(2017, 9 - 1, 12, 12, 24, 42);
ret[0].__ROWID = Math.random();
ret[0].Active = true;
ret[0].Code = "LABEX";
ret[0].Date = new Date(2015, 10 - 1, 4, 18, 2, 34);
ret[0].DateFormat = "M/d/yyyy h:mm:ss tt";
ret[0].DateKey = "Trh_DateTime";
ret[0].Enabled = true;
ret[0].Name = "Lab Map";
ret[0].Remark = "";
ret[0].TimeFormat = "";
ret[0].TimeKey = "";
ret[0].Id = 1;
ret[0].Organizationid = 1;
ret[0].Organization = new Object();
ret[0].Organization.ToString = "[redacted]";
ret[0].Organization._ToString = "[redacted]";
ret[0].Organization.__LOADED = new Date(2017, 9 - 1, 12, 12, 24, 42);
ret[0].Organization.__ROWID = Math.random();
ret[0].Organization.Active = true;
ret[0].Organization.Code = "ARZ";
ret[0].Organization.Date = new Date(2015, 10 - 1, 4, 17, 15, 49);
ret[0].Organization.Enabled = true;
ret[0].Organization.Name = "[redacted]";
ret[0].Organization.Order = 1;
ret[0].Organization.Id = 1;
ret[0].Organization.DataSources = new Array();
```

Figure 0.4: API Response as POJO

The content variation can be extended to new content-types as needed, but it is found that the three types supported by the framework are the most common content types processed by web clients. The example above is from a retrieval operation, where there are no server-side updates, but it is also the case for other server-side update operations, where the result would represent a confirmation of the update or a representation of the created object.

The request payload content for the above example is shown in the figure below.

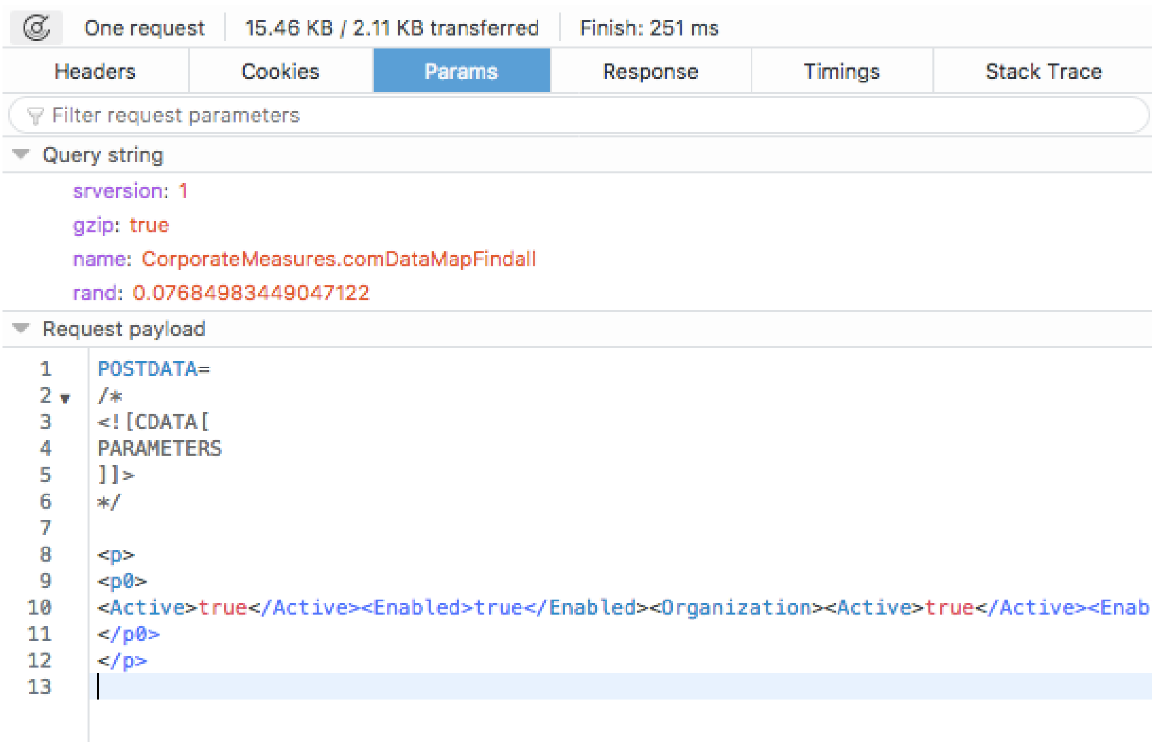

Figure 0.5: Browser HTTP POST request example

The example POST request shows the payload passing as the first parameter the

In addition to the availability of HTTP endpoints, the framework provides a programmable interface library that is delivered to client applications in the form of JavaScript class libraries. To make it clear regarding the structure that is sent to the server in the payload, the same request body is shown in a formatted view:

```xml
<?xml version="1.0" encoding="UTF-8"?>
<p>
   <p0>
      <Active>true</Active>
      <Enabled>true</Enabled>
      <Organization>
         <Active>true</Active>
         <Enabled>true</Enabled>
      </Organization>
   </p0>
</p>
```

Figure 0.6: The formatted XML POST payload

The example post above was generated using such libraries, and This thesis present below the actual JavaScript code that produces the above call:

```javascript
sr._("comDataMapFindall", function(ret) {
    console.log(ret);
    window.DForm.busy(false);
}, {
    Active: true,
    Enabled: true,
    Organization: {
        Active: true,
        Enabled: true,
    }
});
```

Figure 0.7: Client-side example of an API call

This simplified form of a client-side API call is asynchronous, which is the most preferable interaction method for web-based client applications. The library utilizes call-back functions to capture the returned results and forwards the parsed content back to these functions. The same content that is returned in the response of this call is now shown from within the browser response section of the call.

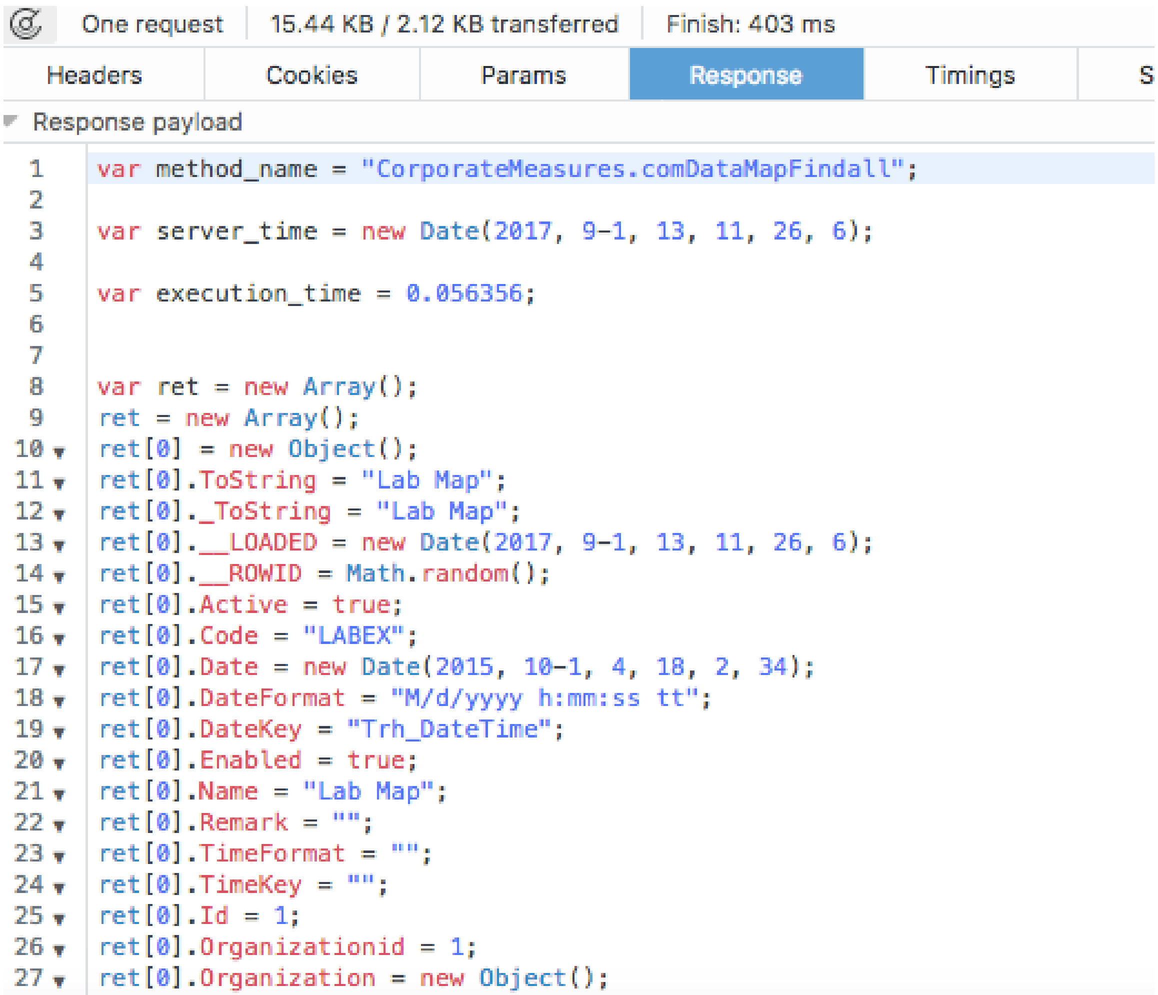

Figure 0.8: Sample response from a browser API call

The default response-type for the API calls is a POJO (Plain Old JavaScript Object) as shown in the above figure, which is easily consumed by most web-based client applications

The DataMap class needs to be defined by the loading process owner, so the following information is created in the system:

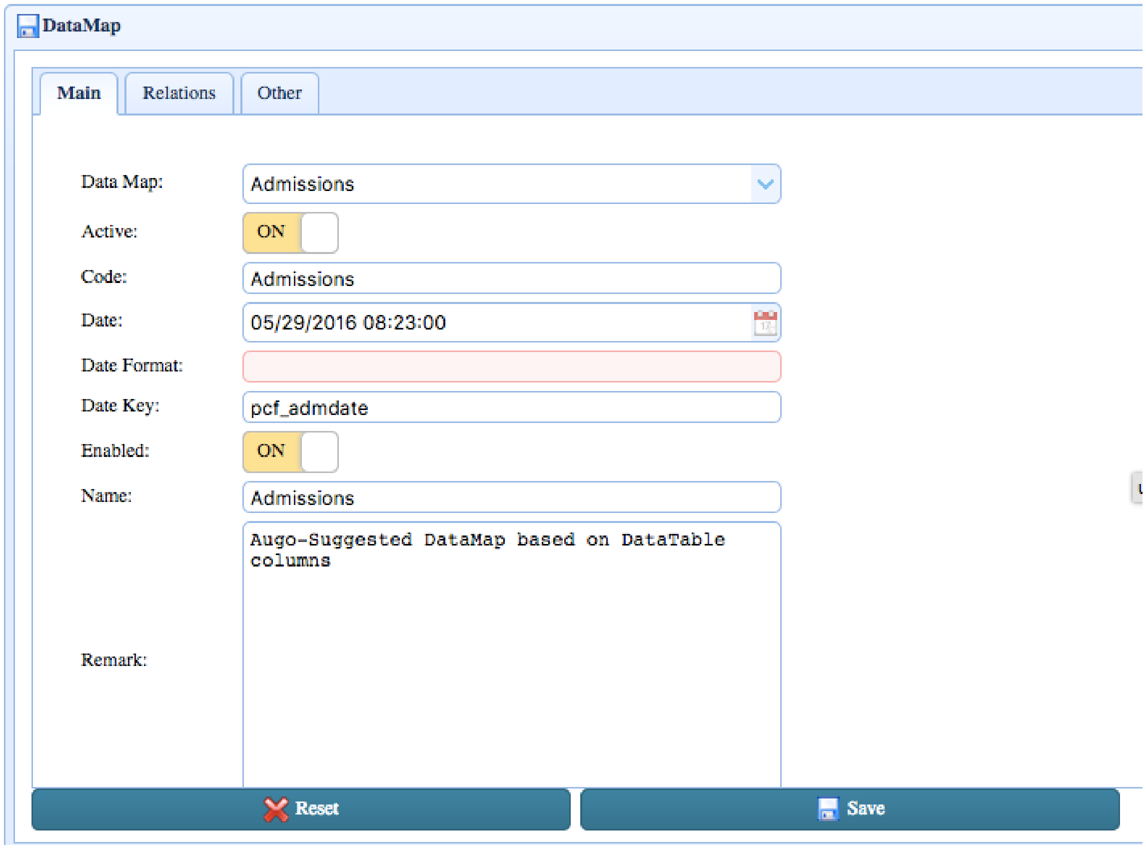

Figure 0.9: The first page DataMap definition widget

The "Date Key" field decides which column of the tabular import-set is used to track the date-stamp for every performance transaction. The details of the data mapping are defined in two tables, the DataField and the DataMapValue. The first defines the tabular column pairs that reference to an Identity in the system.

Appendix B

The following appendix explains the details of the framework system's usage by the users and how they used it to achieve the tasks required from them. The appendix contains many screenshots from the system's usage and describes a use-case from one of the users.

The user is asked to contribute on monthly basis the patient admission aggregations, where the aggregations would be distributed based on the following criteria:

Distribution	Groups	Date Range	Breakdown	Data Sets
Admissions by Type	Patient Type	Yearly	Monthly	Admissions Status
Inpatients by Referrals	Patient Type, Physician	Yearly	Monthly	Admissions Status
Outpatients by Referrals	Patient Type, Physician	Yearly	Monthly	Admissions Status
Inpatients by Attending	Patient Type, Physician	Yearly	Monthly	Admissions Status
Outpatients by Attending	Patient Type, Physician	Yearly	Monthly	Admissions Status
Inpatients by Department	Patient Type, Department	Yearly	Monthly	Admissions Status
Outpatients by Department	Patient Type, Department	Yearly	Monthly	Admissions Status
Inpatients by Readmission	Patient Type, Admitted Before	Yearly	Monthly	Admissions Status
Inpatients by Coverage	Patient Type, Coverage	Yearly	Monthly	Admissions Status
Outpatients by Coverage	Patient Type, Coverage	Yearly	Monthly	Admissions Status

Table 0.1: Admission aggregations required for the monthly status report

The task of the user is divided into three steps. The first stage of the user tasks is to define the data model that will govern the incoming data from the admissions information system. The user is familiar with this data, as she uses this data from the information system using excel exports and knows the structure of this output file. For every entry in Table 0.1 the user is expected to create a column chart spanning a yearly performance of admission data, aggregated by the "Groups" column in this table and broken down over a monthly period. The user would need information from the data set of admissions, which contains similar data to the excel export from the hospital information system.

To define the data model, the user starts with the following screen in the framework system:

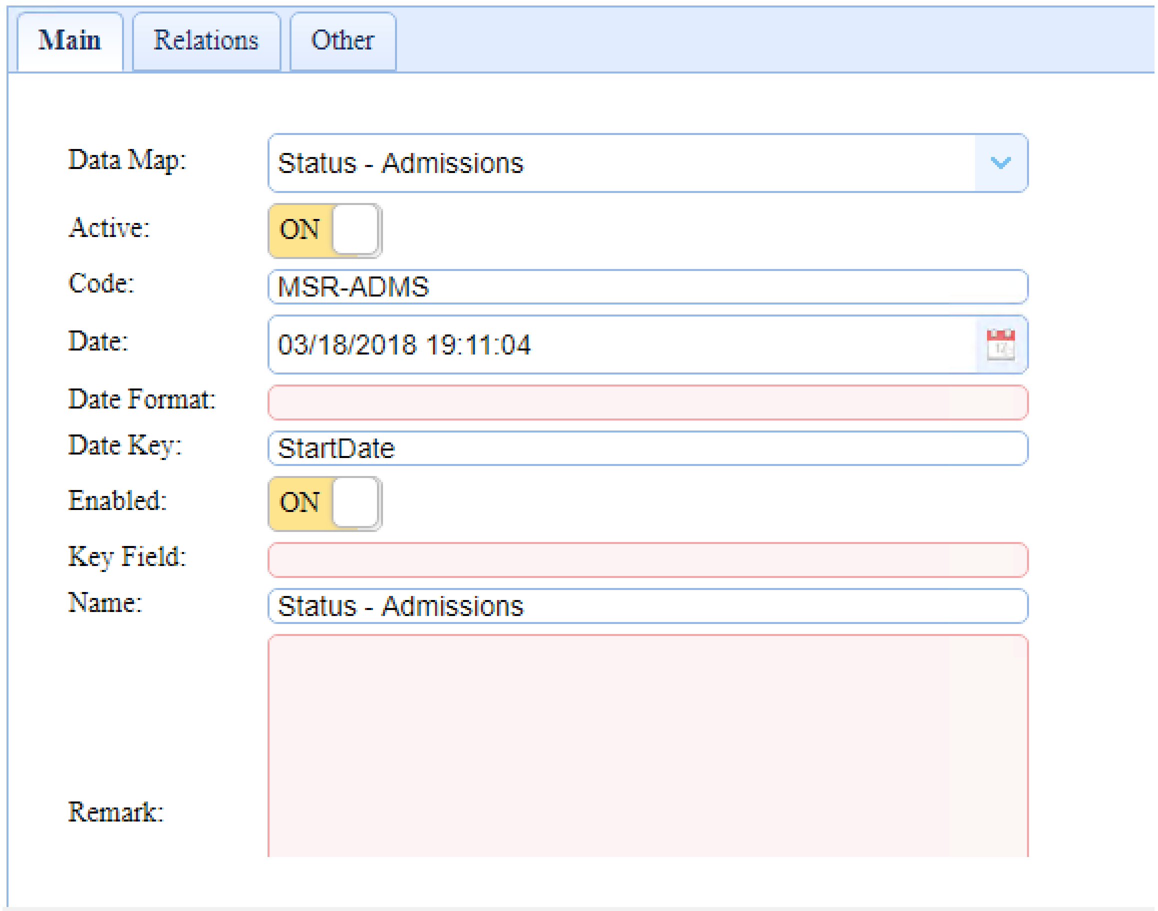

Figure 0.10: Define a data map for the Admission Status

Once the data map is defined, the user needs to define the data fields that are related to this data map. An example is the “Admitted Before” condition, and this is defined in the following two-tab screen:

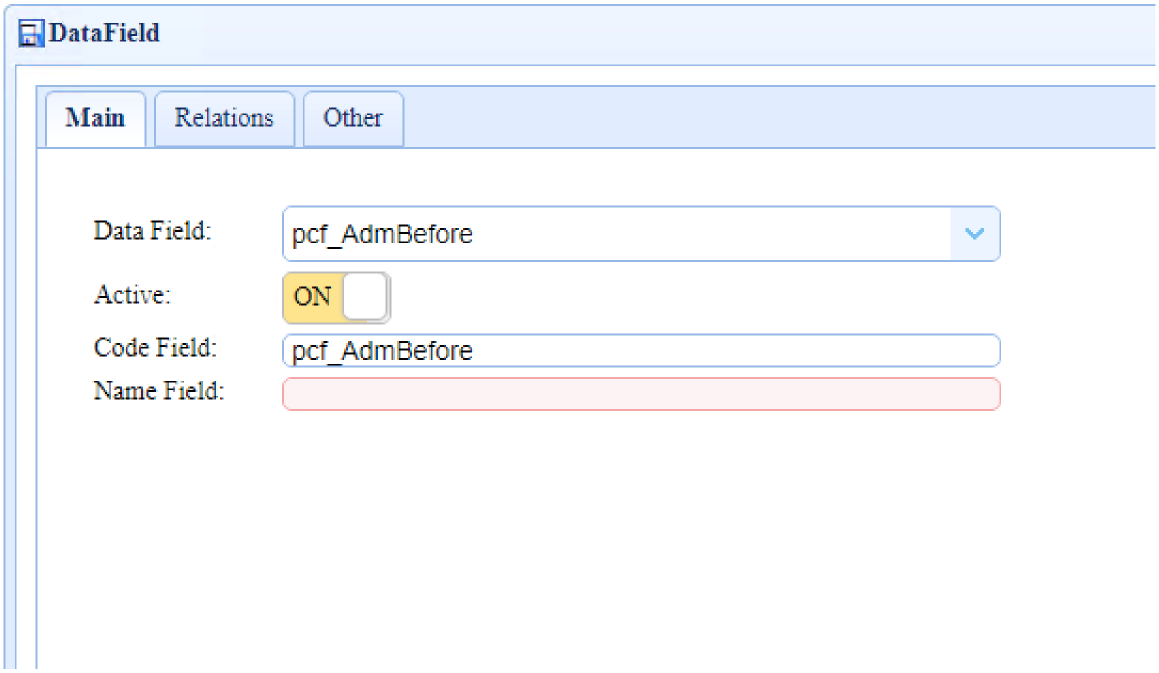

Figure 0.11: Data Field definition screen for the "Admitted Before" data field - screen 1

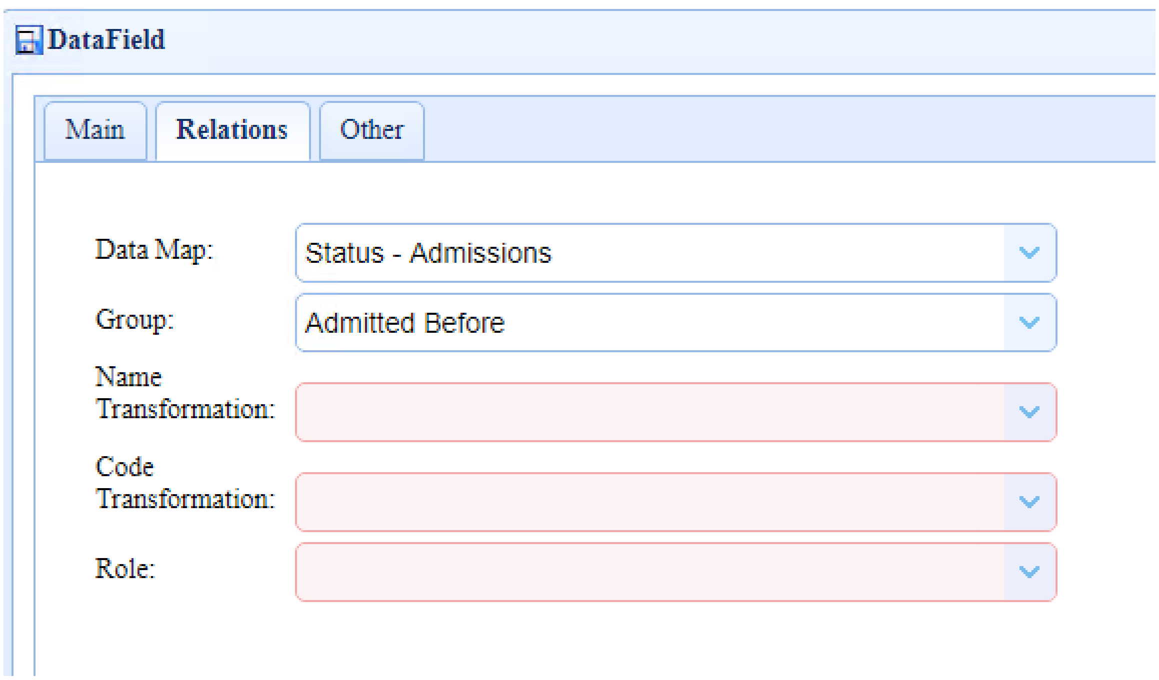

Figure 0.12: Data Field definition screen for the "Admitted Before" data field - screen 2

The screens show how the data field is defined based on the names of the columns in the export data set, and then related to the data map of the “Status – Admissions” defined in the previous step, and the group of identities “Admitted Before” which could either be “Y” for Yes and “N” for No.

The user then defines any value-based measurements that come with the data. In this case where only admission records are loaded, there is no quantitative performance measurements to be made. In other cases, it is possible to have one or more such measurement values, and for that the use might use this screen to do this:

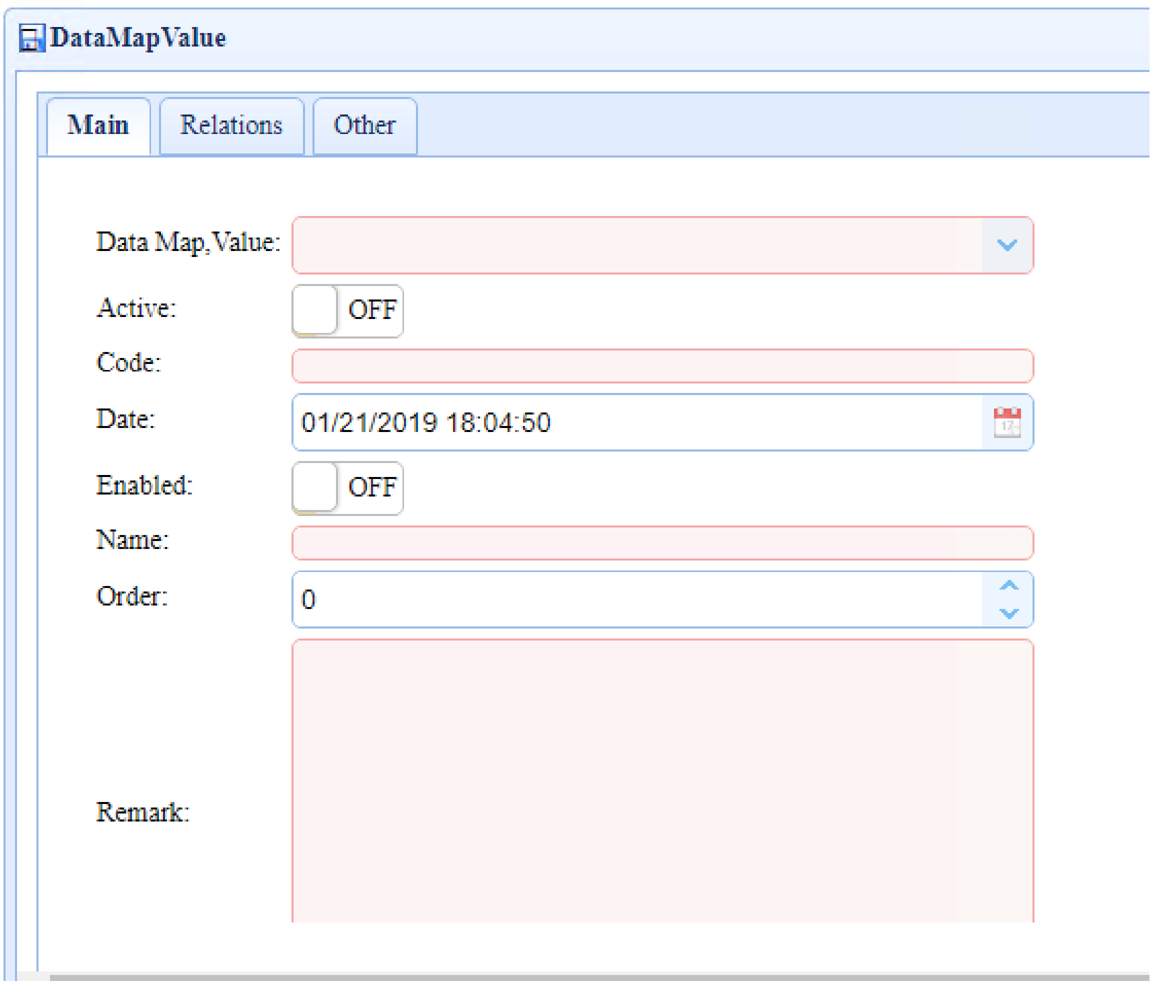

Figure 0.13: Data Map Value setup page - tab 1

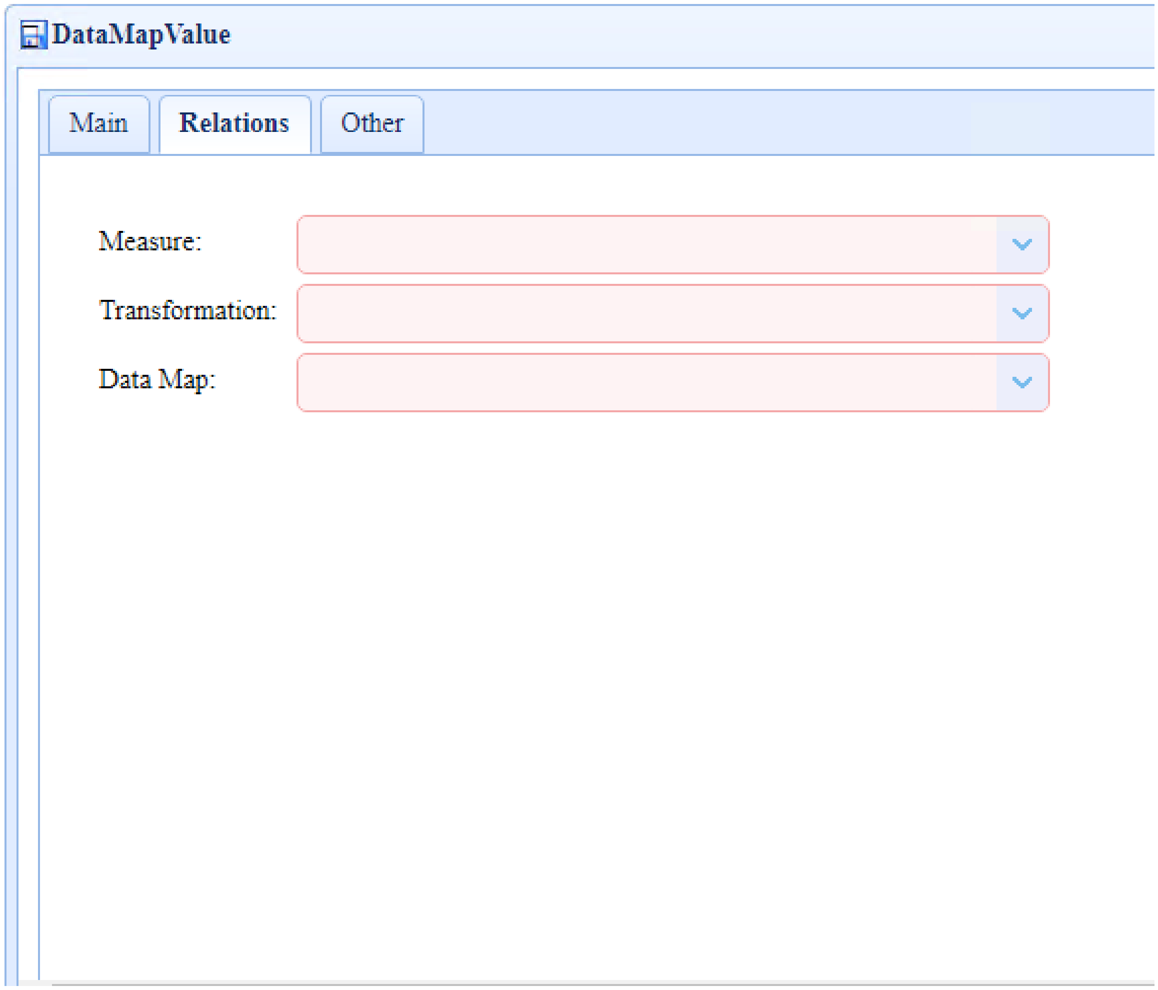

Figure 0.14: Data Map Value setup page - tab 2

The second stage of the preparation for the aggregation tasks in Table 0.1 is to prepare the aggregation mechanism by which the framework should create the results. This is achieved by defining a record in the GroupView table to define the aggregation:

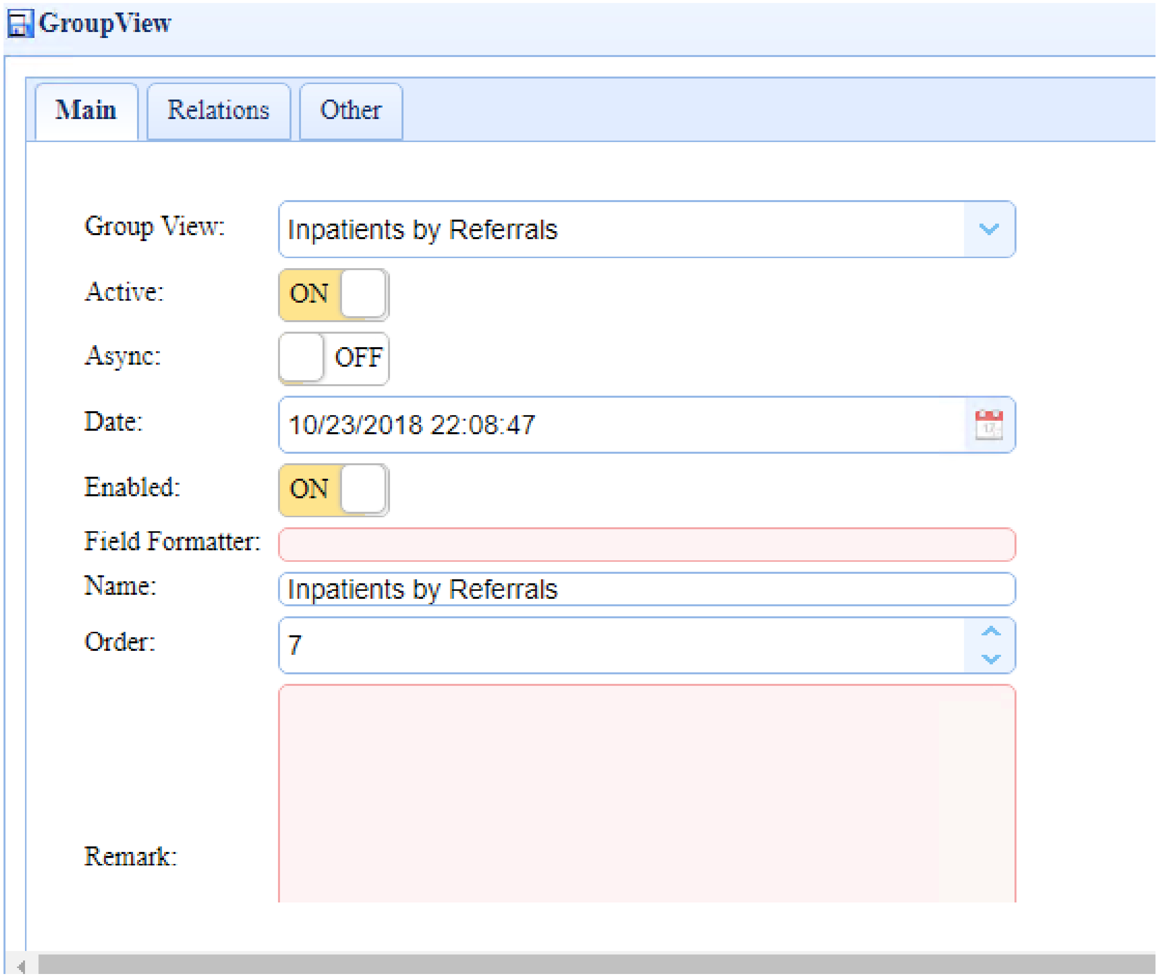

Figure 0.15: Definition screen for the GroupView of the "Inpatients by Referrals" aggregation – tab 1

The user also defines how this group-view should work with the data sets in order to achieve the desired result. The second tab, shown below, instructs the framework system that this group-view should limit the patients to only "Inpatients", should use the "referring doctor" data field defined before as a grouping field, and should use the "count" aggregation to count all the matching patients.

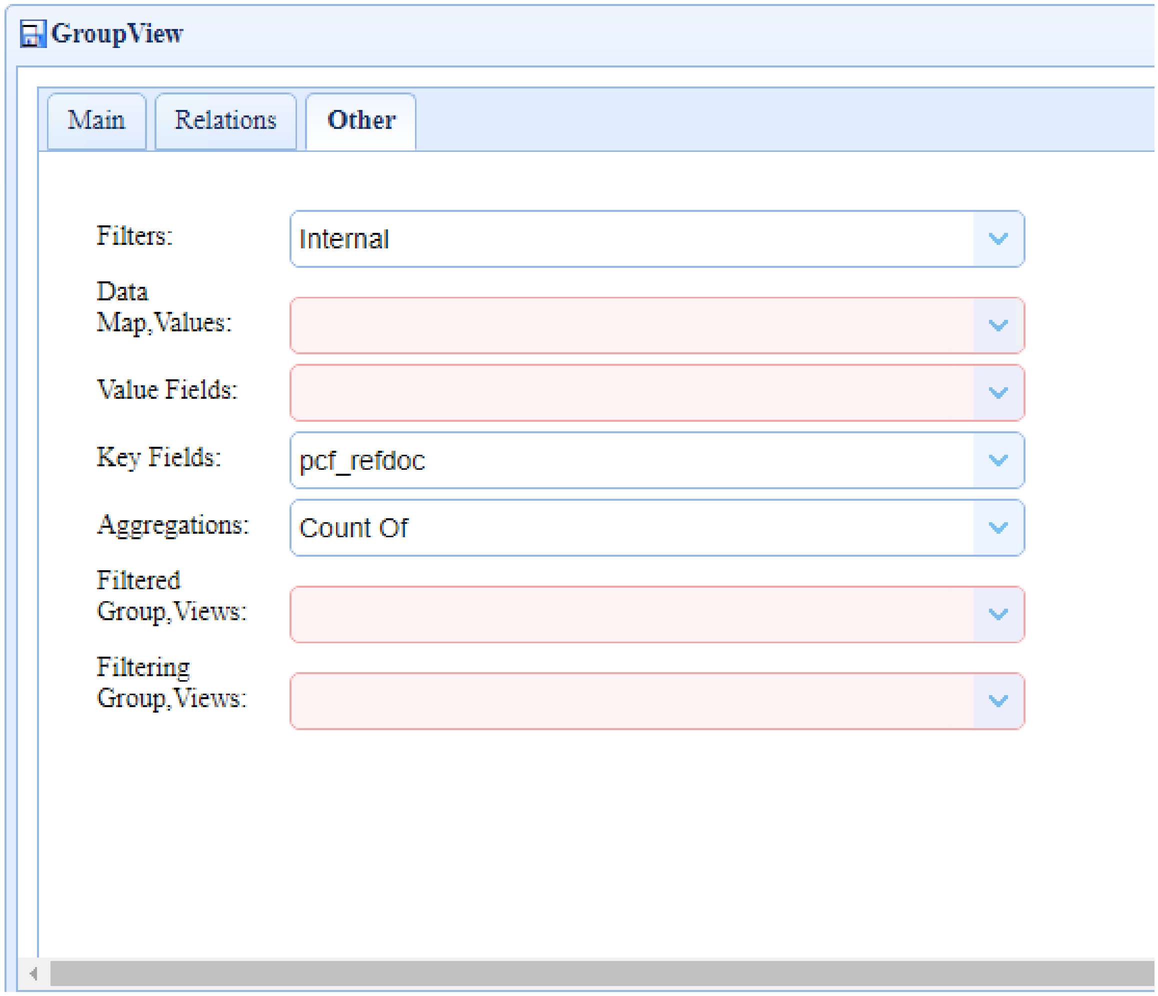

Figure 0.16: Definition screen for the GroupView of the "Inpatients by Referrals" aggregation – tab 2

This concludes the second stage of the tasks required by the user to fulfill the requirements of the monthly status report for the grouping of "Inpatients by Referrals". The third stage is to test the execution of this analytic by specifying other parameters such as the duration and how it should be broken down. The user uses the "Run Analytic" form to test the aggregation analytic. In this example, the user is running the defined analytic using the year 2017 as a date range, over a monthly interval:

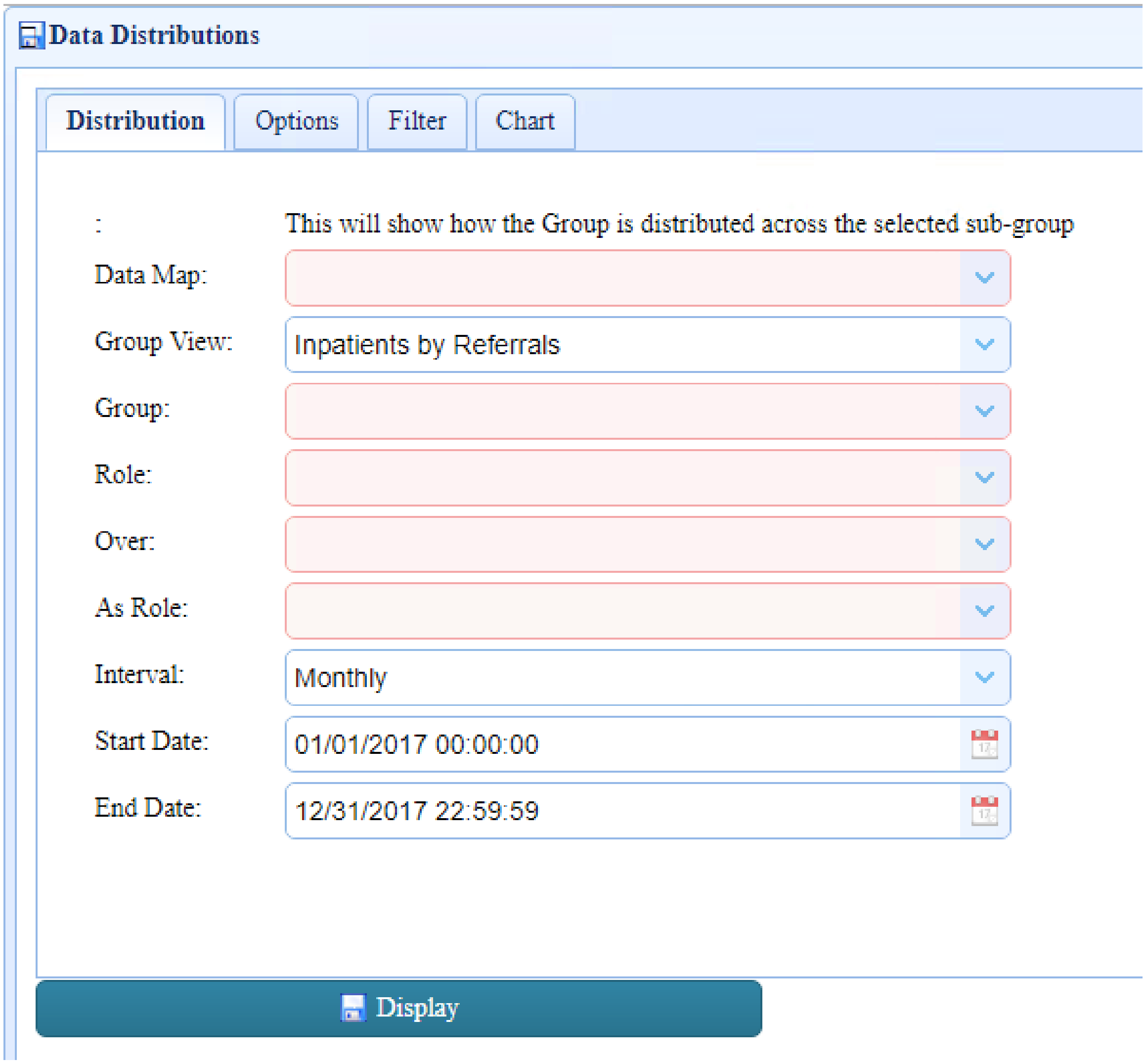

Figure 0.17: Testing the group view against the loaded data

Figure 0.18: The result of the group view defined in the previous steps

Once the user tests that the desired analytic aggregation is correctly representing his data, it is then used in the monthly status report by setting its "Enabled" property to true. It then shows immediately in the monthly status report portal:

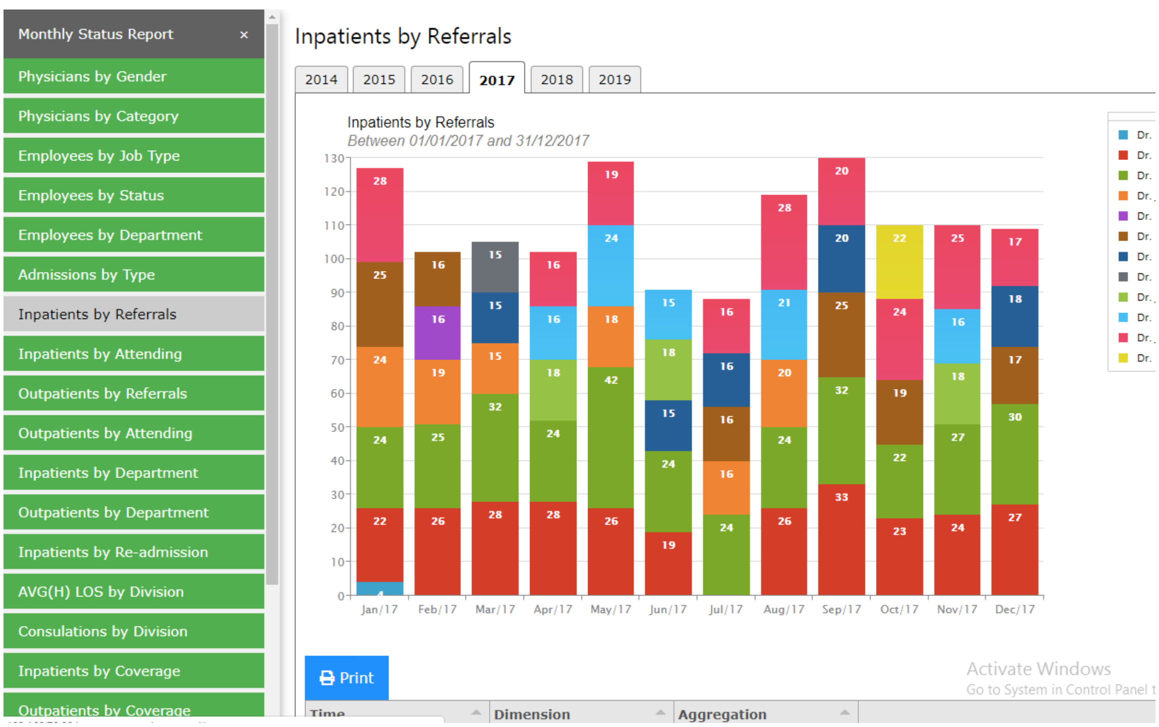

Figure 0.19: The group view is automatically loaded into the monthly status report