UNIVERSITÀ
DEGLI STUDI
DELL'AQUILA

Health Informatics and Data Collaboration (HIDC)

Submitted to:

Prof. Vittorio Cortellessa & Prof.Daniele Di Pompeo

Department of Computer Science and Engineering, and Mathematics

University of L'Aquila (Italy)

Submitted by:

Adam Bouafia

Matricula : 293137

E-mail adam.bouafia@student.univaq.it

Table of contents

1. INTRODUCTION

1.1 Purpose of the Project:

The HIDC system aims to foster robust collaboration among health data analysts by providing a centralized platform for accessing, analyzing, and sharing clinical data. Our objective is to utilize the capabilities of SoBigData++ to establish a health-focused analytical environment. This will enable stakeholders to conduct research, exchange insights, and enhance the collective understanding of health data. A key feature of this system is the seamless integration of data analytics tools, secure communication channels, and compliance dashboards for effective health data management.

1.2 Assumptions of the Project:

1. We will build upon the existing SoBigData++ infrastructure, adapting it as necessary to meet healthcare regulations and address privacy concerns.
2. The system will offer streamlined access to clinical datasets, allowing them to be analyzed and modified within the designated analytical environment, specifically tailored for healthcare research.
3. It will include event notifications and alert management as core elements of the user experience, ensuring timely updates on health data-related activities.
4. Operation will comply with data privacy and security regulations, such as GDPR and HIPAA, under the assumption that SoBigData++ has already established the necessary infrastructure for compliance.
5. The system anticipates differentiated access and functional capabilities for health data analysts and managers, designed to suit their specific roles and interactions within the platform. Furthermore, we expect to utilize the existing database storage and cloud storage resources already available through SoBigData++.

1.3 Scope of the Project:

The project's scope includes a detailed overview of functionalities and user interactions within the HIDC system, illustrated through comprehensive UML diagrams. These diagrams are essential for mapping out the user experience, defining the system architecture, and establishing the workflow for managing health data, fostering collaboration, and ensuring compliance. Our documentation will feature:

- Use case diagrams that outline the responsibilities and interactions of health data analysts and managers.
- Sequence diagrams that depict key user interactions, such as browsing datasets, downloading clinical data, and contributing research findings.

- Component and deployment diagrams that illustrate the system's backend architecture, providing insights into how various elements of the system are integrated and maintained.

1.4 Definitions, Acronyms, and Abbreviations:

- SoBigData++: A community for Social Mining and Big Data Analytics
- UML: Unified Modeling Language
- JAR: Java Archive
- HIDC: Health informatics ans Data Collaboration
- HIPAA: Health Insurance Portability and Accountability Act
- GDPR: general data protection regulation

1.5 References:

1. http://www.sobigdata.eu/int_communication
2. https://plusplus.sobigdata.eu/objectives/
3. https://www.lucidchart.com/pages/uml-deployment-diagram
4. https://www.lucidchart.com/pages/uml-component-diagram

1.6 Overview of the Report:

This report thoroughly documents the design of the system through UML diagrams, providing both visual representations and detailed explanations of the design decisions made.

Here's what each type of diagram represents:

- **Use Case Diagrams:** These diagrams illustrate the interactions between users—namely Health Data Analysts and Health Data Managers—and the system's functions. They help in understanding how different user roles engage with the system.

- **Sequence Diagrams:** These diagrams outline key processes such as browsing datasets, communicating with health professionals, and providing feedback on studies. They detail the step-by-step interactions between system components and users over time.

- **Component Diagrams:** These diagrams detail the various components within the system, outlining their responsibilities and how they interact with one another. This helps in visualizing the system's structure and the flow of information.

- **Deployment Diagrams:** These diagrams show how software artifacts are physically deployed on hardware components, with a focus on maintaining security and compliance within the healthcare context. They are crucial for understanding the system's infrastructure and operational environment.

- **Profile Diagrams:** These diagrams describe the stereotypes, tagged values, and constraints applied to elements within the system, reflecting the specific standards and requirements of the healthcare industry. They are essential for ensuring that the system meets industry-specific regulations and practices.

2. SYSTEM OVERVIEW

2.1 Business Context:

The HIDC system is strategically designed to bridge health informatics with collaborative data analysis. Its main goal is to expand the capabilities of the SoBigData++ infrastructure specifically for healthcare research. This initiative promotes an interdisciplinary approach, encouraging deeper and more insightful explorations into health data analytics, which could lead to significant advancements in healthcare practices and policies.

2.2 System Objectives:

The primary objectives of the HIDC system are as follows:

- **Foster Collaboration:** Create an environment that enables health data analysts to engage in pioneering research and analysis, encouraging collaboration across different specialties and fields.
- **Secure and Compliant Platform:** Provide a secure platform for sharing, discussing, and advancing knowledge within health informatics, ensuring compliance with all relevant regulations and privacy standards.
- **Integration of Tools:** Seamlessly integrate health data analytics tools with the SoBigData++ e-infrastructure, offering a bespoke experience that enhances health data exploration and contribution. This integration aims to make sophisticated data analysis tools more accessible to health researchers, thereby enriching the quality and impact of their work.

2.3 High-Level System Description:

The HIDC platform is a sophisticated web-based portal specifically designed for the in-depth exploration and analysis of clinical data. It enables health data analysts to access diverse datasets, utilize advanced analytical tools, and participate in collaborative research projects. Key features of the platform include secure messaging for privacy-ensured communication, compliance dashboards that help users adhere to various regulations, and robust data publication systems that facilitate the sharing of findings. All these functionalities are built upon the reliable and scalable foundation provided by SoBigData++.

3. METHODOLOGY

3.1 Overview of Approach:

Our development strategy is methodically structured, focusing on the creation of UML diagrams step-by-step. This approach captures the system's multifaceted interactions and architectural design, ensuring that we represent all functionalities comprehensively. These diagrams serve as a blueprint for understanding the system’s structure and workflow, guiding the development process and facilitating communication among team members.

3.2 Tools Used:

For our UML diagramming and model validation, we have chosen MagicDraw. This tool is renowned for its robust features that support a high level of design precision and efficiency. MagicDraw helps in creating accurate and detailed representations of the system architecture, which is crucial for successful implementation and future maintenance.

3.3 Limitations and Assumptions:

Our project operates under several key assumptions. Firstly, it is presumed that users have a basic understanding of health data systems, enabling them to effectively navigate and utilize the HIDC platform. Additionally, user access to the SoBigData++ resources is assumed to be seamless, facilitating an uninterrupted experience. The scope of the project currently aligns with the existing capabilities of SoBigData++, although we have made provisions for future scalability. This includes the potential integration of additional health data sources and analytical tools as they become available and as the need within the healthcare research community grows.

4. UML MODELING

In this section, we present a series of comprehensive UML diagrams that outline the system architecture, depict user interactions, and illustrate the data flow within the HIDC system. These diagrams are crucial as they serve as the blueprint for the system's development and implementation, ensuring that all aspects of the system are thoroughly planned and understood.

4.1. Use Case Diagram:

The Use Case Diagram offers a clear visualization of how the HIDC system interacts with its primary users: the Health Data Analyst and the Health Data Manager. This diagram highlights the core functionalities available to these users, including browsing health datasets, downloading clinical data, contributing research findings, and communicating with health professionals. The diagram serves as a foundational tool in understanding the interactions and processes that define the user experience within the HIDC system.

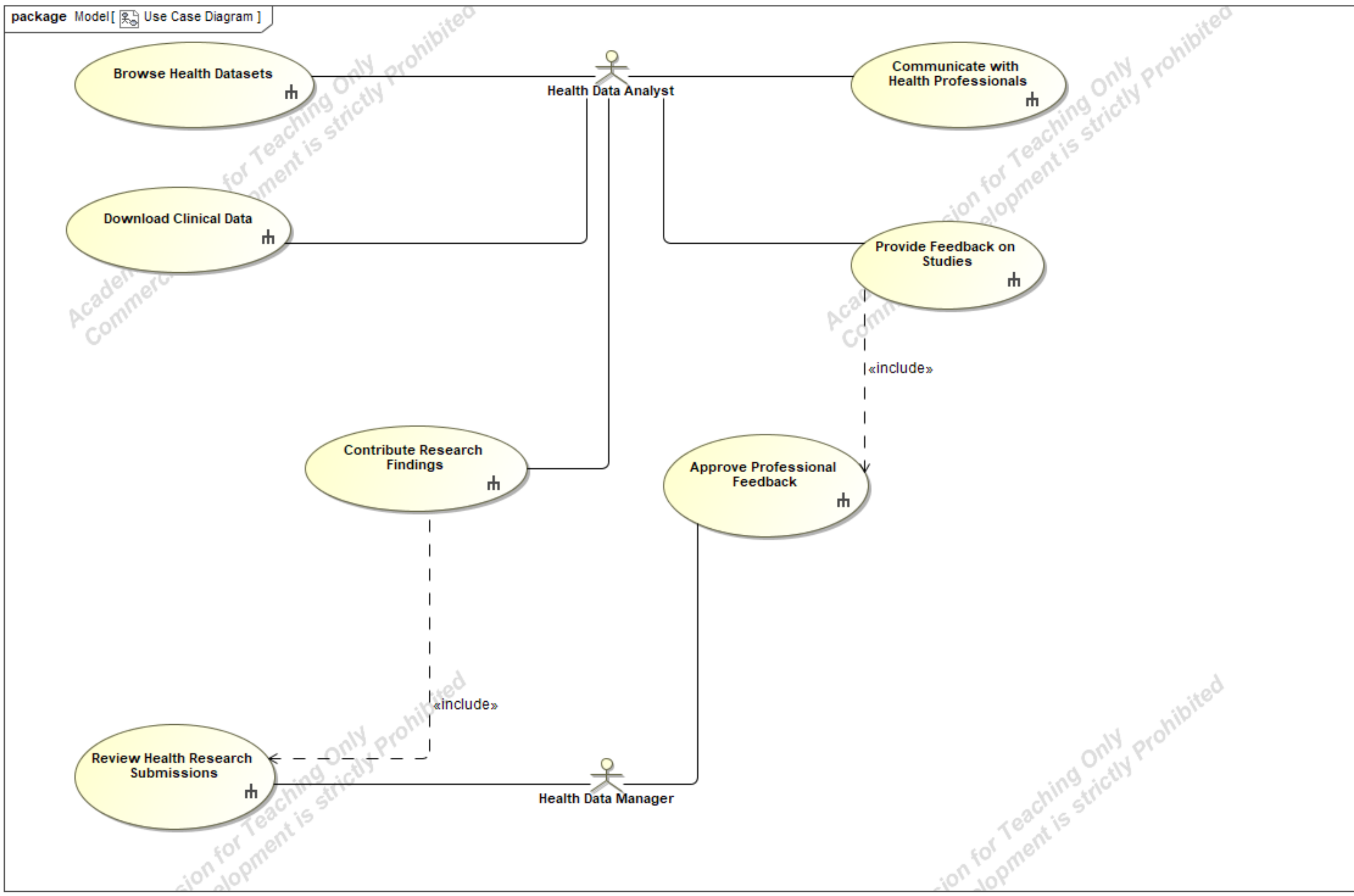

Fig 1: use case diagram

4.1.1 Actors Identification

In the HIDC system, two main actors interact with the platform:

Health Data Analyst: These individuals engage extensively with the HIDC system, accessing datasets, contributing findings, and collaborating with peers. They are pivotal in driving forward the research and analytical work within the system.

Health Data Manager: These users, often serving as health data custodians or administrators, manage the datasets, approve access, and handle feedback on studies. Their role ensures the integrity and regulatory compliance of the data management process.

4.1.2 Use Case Scenarios

1. **Browse Health Datasets (Health Data Analyst):** Analysts access a diverse array of datasets to identify relevant data for their research needs.
2. **Download Clinical Data (Health Data Analyst):** Once relevant datasets are identified, analysts can download the data for more in-depth analysis.
3. **Contribute Research Findings (Health Data Analyst):** Analysts submit their research findings, which may include new insights or datasets derived from their investigations.
4. **Communicate with Health Professionals (Health Data Analyst):** Analysts engage with health professionals to discuss research findings, seek expert opinions, or collaborate on further studies.
5. **Provide Feedback on Studies (Health Data Manager):** Managers review feedback from peers on the published studies or data used, facilitating a cycle of continuous improvement.
6. **Approve Professional Feedback (Health Data Manager):** Managers assess and validate feedback provided by professionals, ensuring the accuracy and relevance of the data or research findings before wide dissemination.
7. **Review Health Research Submissions (Health Data Manager):** Managers evaluate new research contributions to ensure they meet quality standards, relevance, and compliance with data use agreements.

4.2. Profile Diagram:

The profile diagram is essential in laying out the foundational architectural design of the HIDC system by defining its key building blocks and their classifications. It ensures consistency and clarity across the system’s structure, serving as a guiding framework for the development and deployment of the HIDC platform.

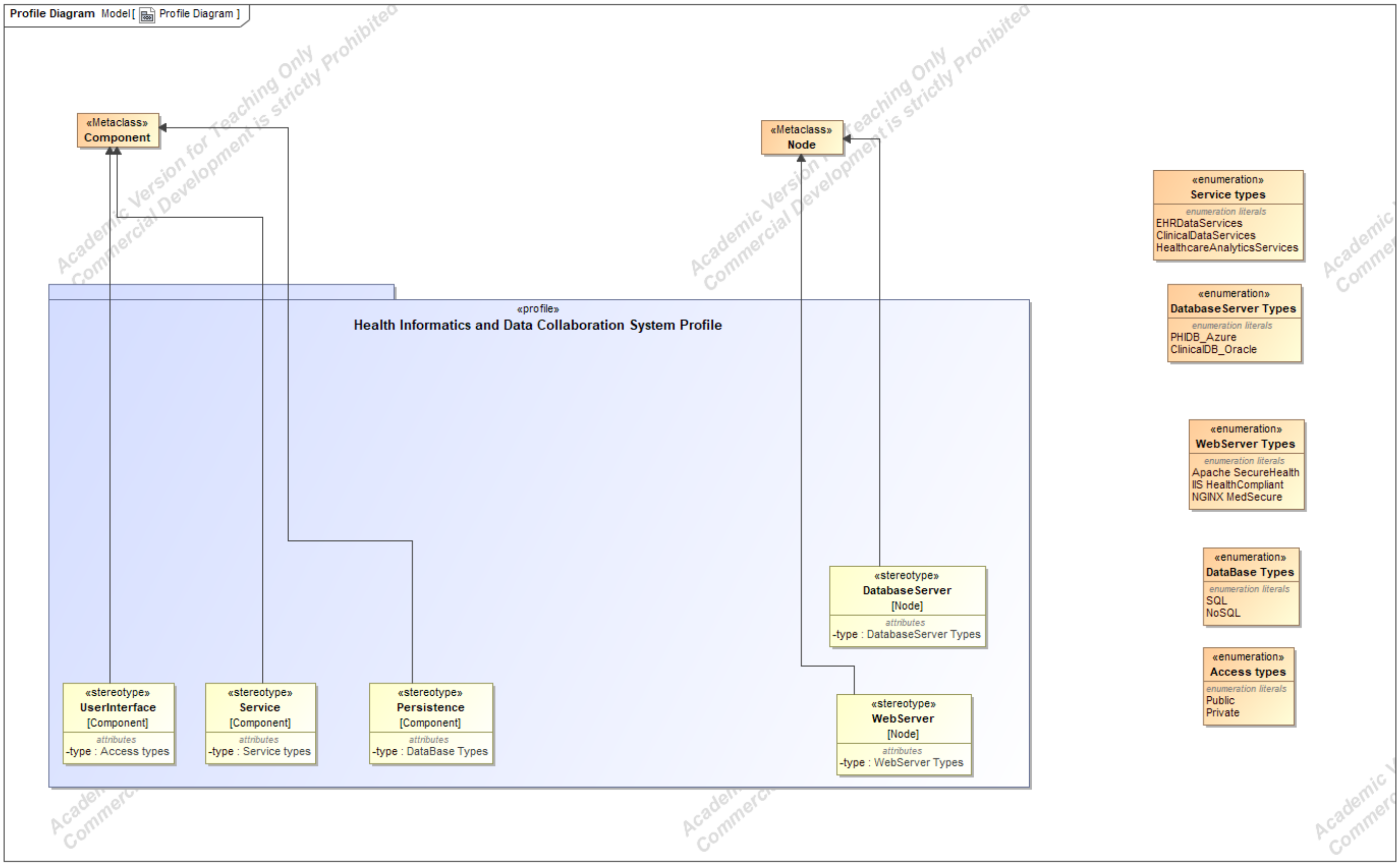

Fig 2: Profile Diagram

4.2.1. UML Metaclasses:

- **Component**: This metaclass is specialized with stereotypes to denote functional units of the Health Informatics and Data Collaboration System (HIDC). These components are critical as they encapsulate the specific capabilities provided by the system to its users, ensuring functional effectiveness and user-centric operations.

- **Node**: Enhanced with the DatabaseServer and WebServer stereotypes, this metaclass characterizes the infrastructural elements of the HIDC system. It signifies the physical and logical locations where data and services are stored and executed, forming the backbone of the system's infrastructure.

4.2.2. Stereotypes Defined:

- **UserInterface**: Derived from the Component metaclass, this stereotype represents the various user interfaces designed for different stakeholders within the HIDC system. It is categorized by Access Types (Public, Private, Secure), tailoring the system's accessibility to meet specific security needs.
- **Service**: Derived from the Component metaclass, this stereotype represents the various user interfaces designed for different stakeholders within the HIDC system. It is categorized by Access Types (Public, Private, Secure), tailoring the system's accessibility to meet specific security needs.
- **Persistence**: This stereotype, originating from the Component metaclass, encapsulates the data storage and retrieval mechanisms. It is crucial for maintaining the integrity and availability of health data within the HIDC system.
- **DatabaseServer:** Derived from the Node metaclass, this stereotype indicates the database servers that support the HIDC system's data persistence needs. It is classified by Database Server Types, specifying the technology stacks used.
- **WebServer:** Originating from the Node metaclass, this stereotype highlights the web servers that deploy the user interfaces and services of the HIDC system. It is classified by Web Server Types, denoting the specific technologies used.

4.2.3. Enumerations:

- **Service Type**This enumeration categorizes the various types of services within the HIDC system, such as *EHRDataServices*, *ClinicalDataServices*, and *HealthcareAnalyticsServices*. Each category is tailored to a specific service domain.
- **Database Server Types:** Enumerates the types of database servers like *PHIDB_Azure*, *ClinicalDB_Oracle*, etc., defining the technology stacks used for data storage.
- **Web Server Types**: Details the server types like *NGINX MedSecure, IIS HealthCompliant*, and Apache SecureHealth, representing the technologies that host the user interfaces and services.
- **Access Types**: Details the accessibility categories for the UserInterface stereotype, distinguishing between Public, Private, and Secure types based on the required level of accessibility and security for different user interactions.

4.3. Component Diagram:

The component diagram is instrumental in providing a detailed representation of the HIDC system's architecture, illustrating the functional roles of each component within the system and their interaction through well-defined interfaces. Here is a detailed breakdown of the primary components of the HIDC system:

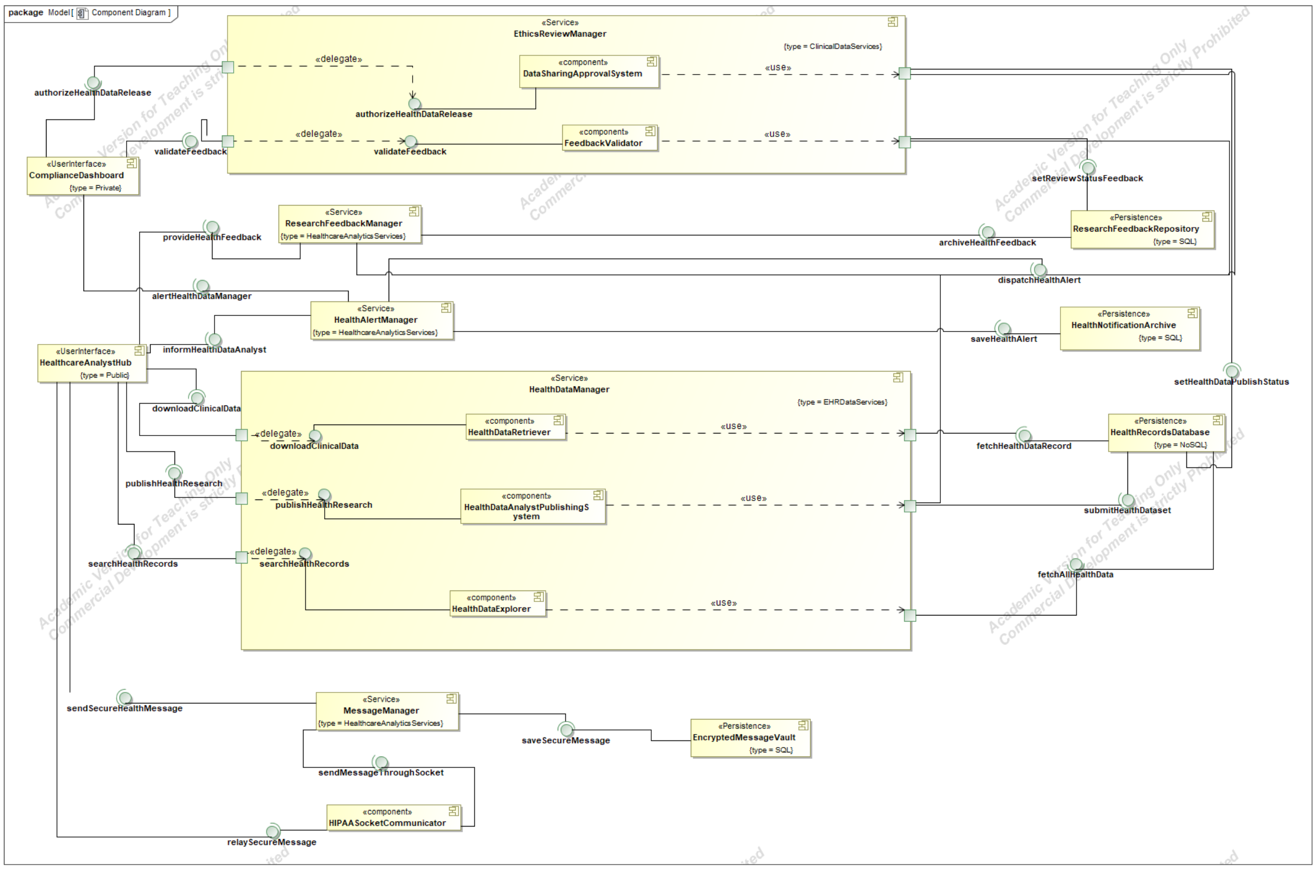

Fig 3: Component diagram

4.3.1. Component Descriptions

1. **HealthcareAnalystHub:** This is the primary user interface component for health data analysts. It facilitates browsing datasets, initiating downloads, and contributing research findings, serving as the central interaction point for analysts.
2. **ComplianceDashboard:** A specialized user interface for health data managers, this component is crucial for overseeing compliance aspects, validating feedback, and managing

data release authorizations. It helps ensure that all data handling adheres to legal and ethical standards.

3. **EthicsReviewManager:** This service component manages the ethical review processes and is interconnected with the **DataSharingApprovalSystem**. It ensures that all health data handling is conducted following established ethical standards and guidelines.
4. **HealthDataManager:** A core service component that orchestrates activities related to health data management. This includes the publication of research findings and managing access to health records, pivotal for maintaining the integrity and confidentiality of the data.
5. **HealthDataRetriever**: Working in tandem with the **HealthDataManager**, this component facilitates the download and retrieval of clinical data from the **HealthRecordsDatabase**. It ensures efficient and secure data access for health data analysts.
6. **HealthDataExplorer:** This auxiliary component enhances the ability to explore and search health records, providing an informative and interactive user experience for researchers and analysts looking to mine data for insights.
7. **MessageManager and HIPAASocketCommunicator:** These components are key for ensuring secure communication within the system. They manage the transmission of secure health messages and interface with encrypted data vaults to safeguard sensitive information.
8. **ResearchFeedbackManager and FeedbackValidator:** These components handle the feedback provided by health data analysts on various studies. They facilitate a comprehensive review process to ensure the accuracy and reliability of the data before it is released to the public.
9. **HealthNotificationArchive:** As a persistence component, it serves as a repository for health notifications and alerts. It plays an essential role in the timely dissemination of crucial health information to relevant stakeholders.
10. **DataSharingApprovalSystem:** This component acts as a control system for data sharing operations. It ensures that data publication and sharing are in line with consented guidelines and comply with regulatory standards.

By understanding the roles and interactions of these components, stakeholders can ensure that the system's architecture is not only robust and integrated but also adaptable to the evolving needs of health data management and analysis. This diagram serves as a blueprint for the system's implementation and ongoing development, highlighting the importance of each component in the overall functionality of the HIDC system.

4.4. Sequence Diagram

The sequence diagrams in the HIDC system provide a dynamic and interactive visual representation of the process flows that occur between Health Data Analysts and Managers within the health informatics ecosystem. These diagrams are critical for mapping out the structured order of messages and interactions among different system components, helping to visualize the functional interactions and data flow.

4.4.1 SEQUENCE DIAGRAM FOR 'BROWSE HEALTH DATASETS'

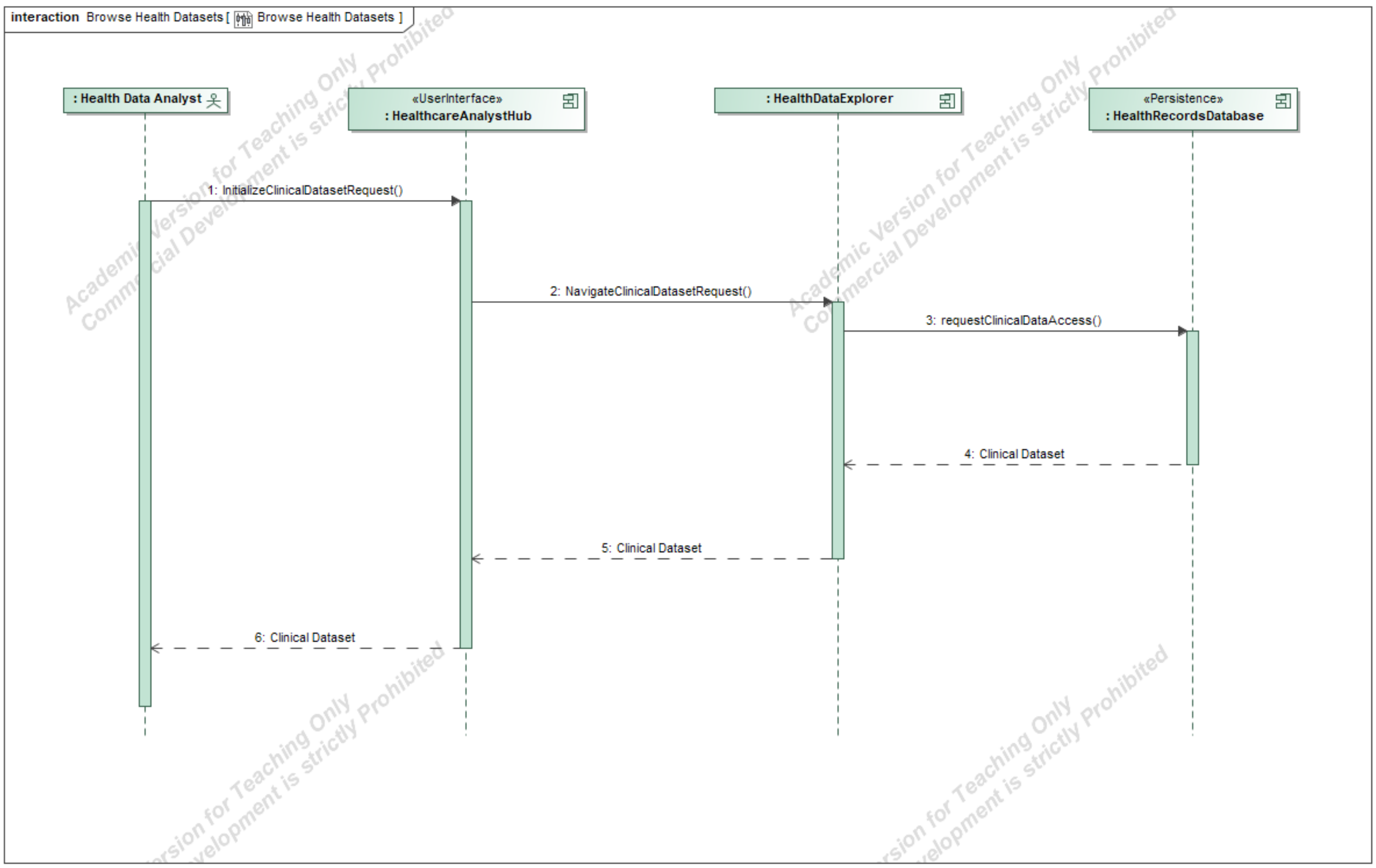

Fig 4: Sequence Diagram for `'Browse Health Datasets` use case

This diagram outlines the steps taken by Health Data Analysts when they initiate a request to browse available health datasets:

- The Health Data Analyst starts by interacting with the **HealthcareAnalystHub** interface.
- The request is then processed by the **HealthDataExplorer** component, which navigates and requests access to clinical data from the **HealthRecordsDatabase**.
- Finally, the requested clinical dataset is retrieved and provided back to the Health Data Analyst, completing the data browsing cycle.

4.4.2 SEQUENCE DIAGRAM FOR 'COMMUNICATE WITH HEALTH PROFESSIONALS'

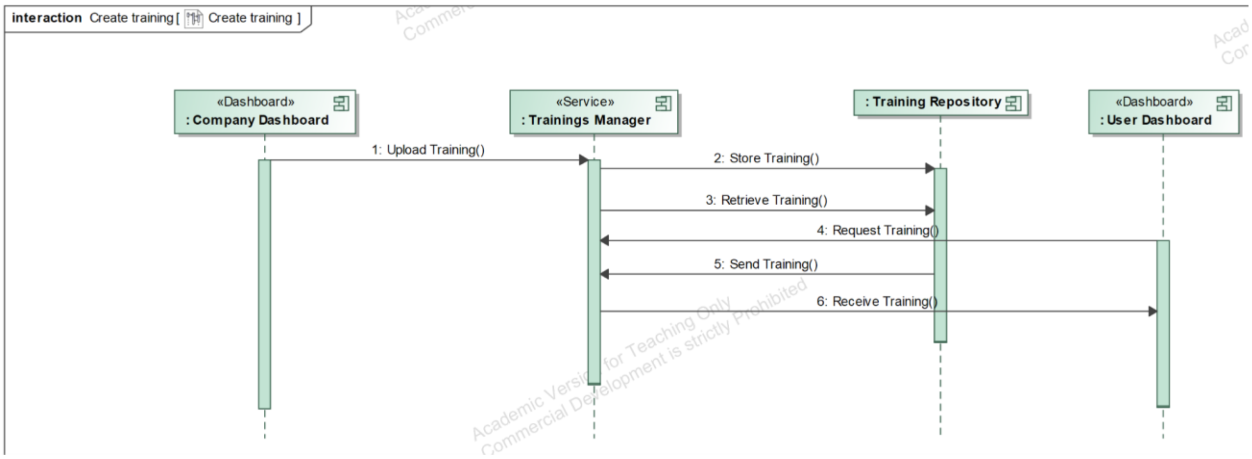

Fig 5: Sequence Diagram for `Communicate with Health Professionals' use case

This diagram captures the secure communication protocol between Health Data Analysts and Health Professionals:

- The Health Data Analyst initiates the creation of a secure message.
- The **MessageManager** processes and relays this message.
- The **EncryptedMessageVault** ensures the data's integrity and confidentiality.
- The **HIPAASocketCommunicator** then facilitates the secure transmission of the message, demonstrating the system's commitment to healthcare compliance and data security.

4.4.3 SEQUENCE DIAGRAM FOR 'DOWNLOAD CLINICAL DATA'

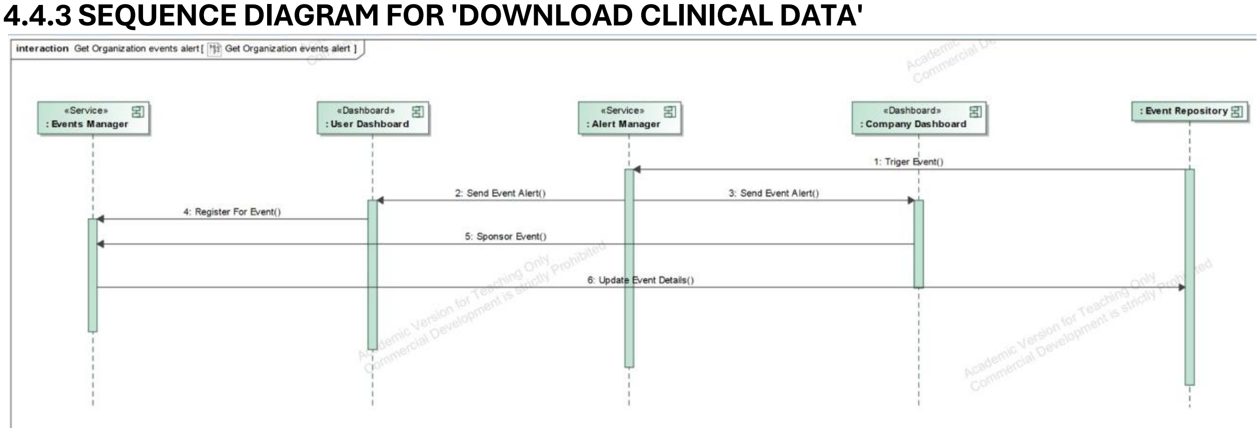

Fig 6: Sequence Diagram for `Download Clinical Data` use case

This diagram details the process for downloading clinical data:

- Initiated through the **HealthcareAnalystHub**, the Health Data Analyst requests a clinical data download.
- The **HealthDataRetriever** interacts with the **HealthRecordsDatabase** to retrieve and deliver the requested data, ensuring a smooth and secure data download process.

4.4.4 Sequence Diagram for 'Contribute Research Findings'

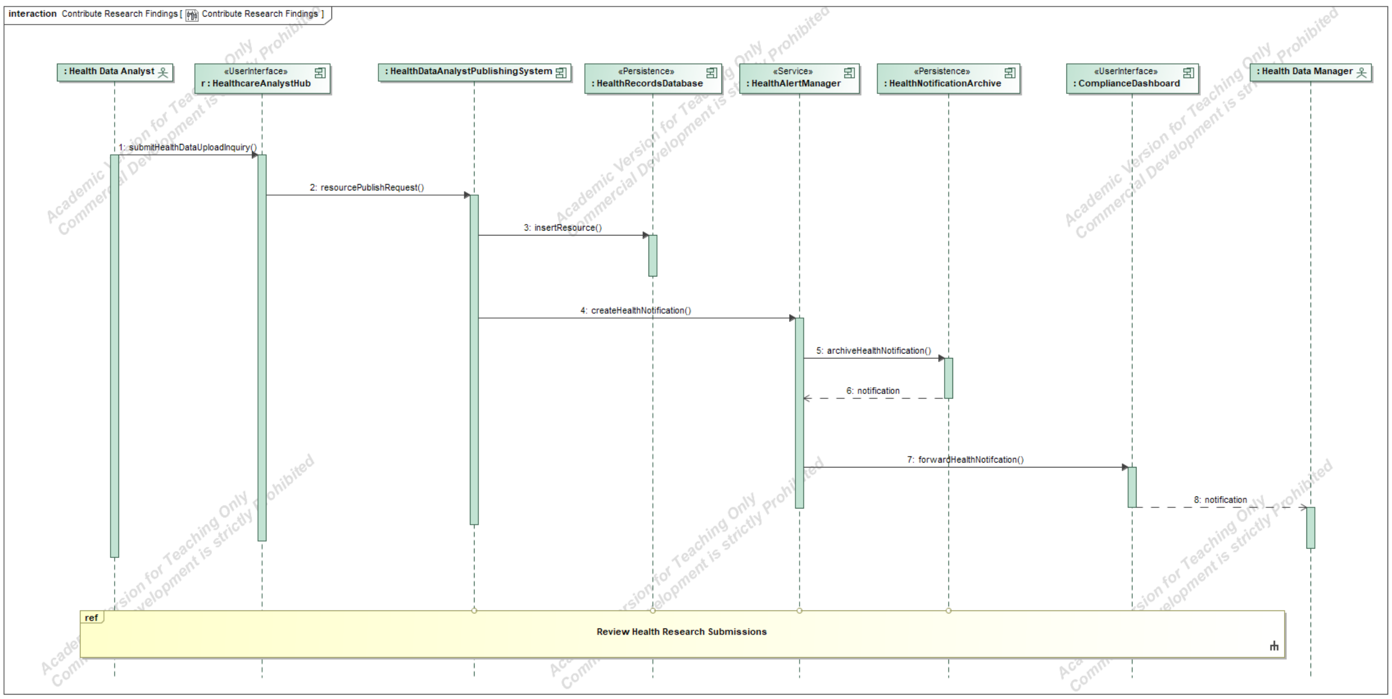

Fig 7: Sequence Diagram for 'Contribute Research Findings' use case

This diagram illustrates the process for submitting research findings:

- The Health Data Analyst initiates a data upload inquiry through the **HealthcareAnalystHub**.
- This triggers the **HealthDataPublishingSystem** to request the insertion of resources into the **HealthRecordsDatabase**.
- The **HealthAlertManager** manages notification and archiving processes, closing the loop with feedback communicated back to the user.

4.4.5 SEQUENCE DIAGRAM FOR 'PROVIDE FEEDBACK ON STUDIES'

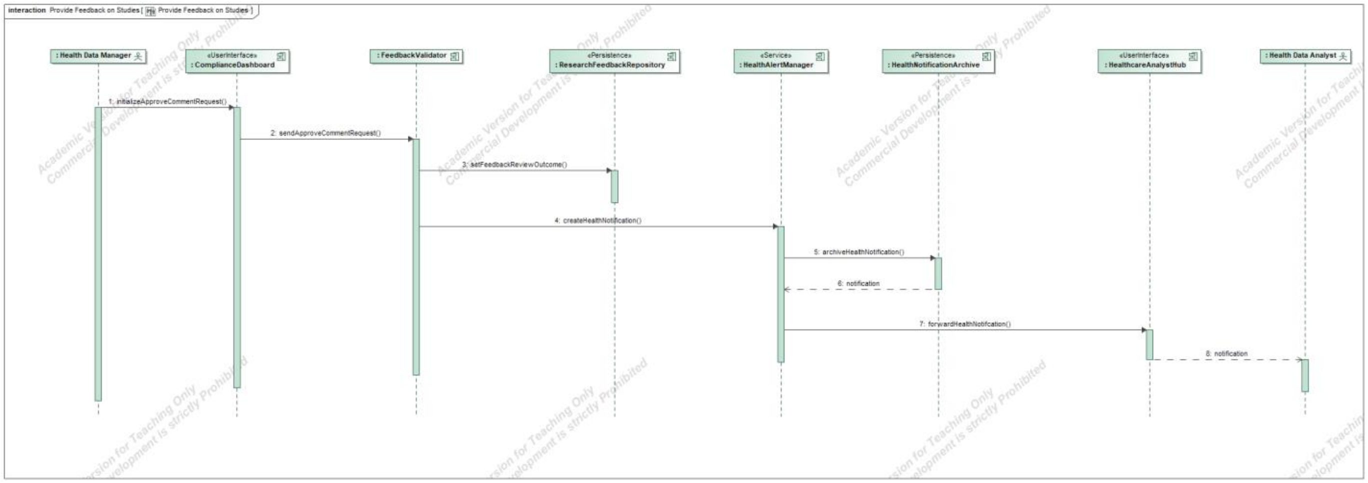

Fig 8: Sequence Diagram for 'Provide Feedback on Studies' use case

This diagram shows how Health Data Managers provide feedback on clinical studies:

- Starting with the **ComplianceDashboard** interface, the process involves validating feedback through the **FeedbackValidator**.
- Updates to records and management of health notifications are handled by the **HealthAlertManager** to efficiently disseminate the information.

4.4.6 SEQUENCE DIAGRAM FOR REVIEW HEALTH RESEARCH SUBMISSIONS'

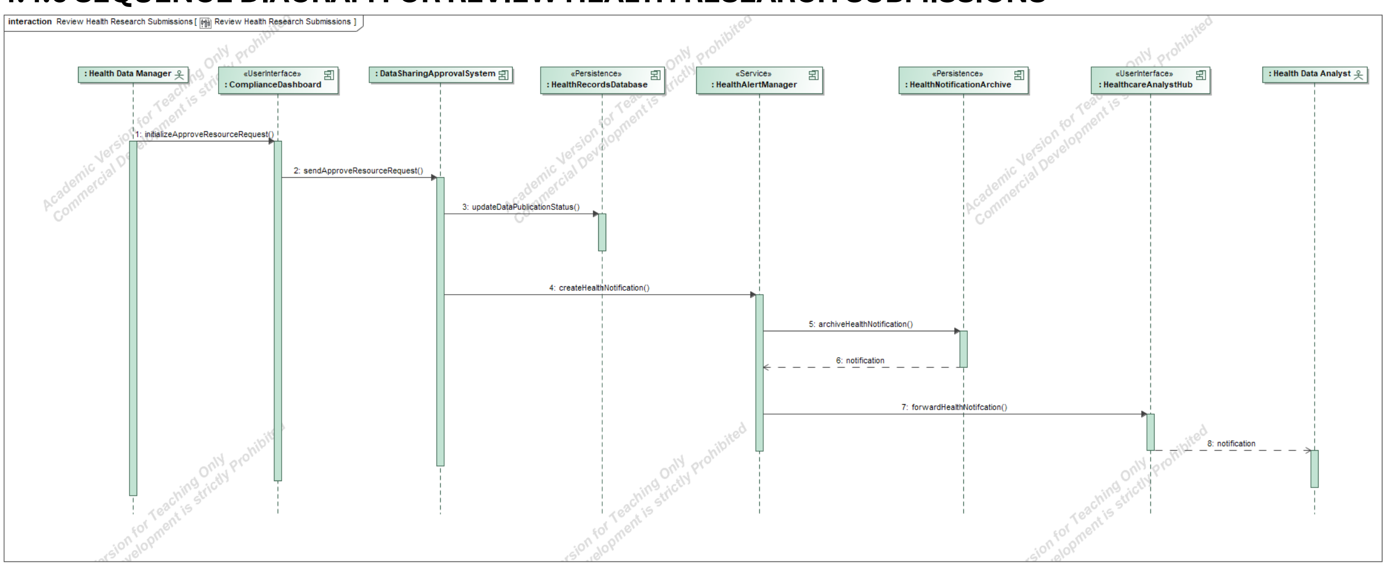

Fig 9: Sequence Diagram for 'Review Health Research Submissions' use case

This sequence details the review of health research submissions:

- Initiated via the **ComplianceDashboard**, the Health Data Manager interfaces with the **DataSharingApprovalSystem** to update publication statuses.

- The flow includes creating notifications and archiving submissions, highlighting the system's comprehensive mechanisms for review.

These sequence diagrams are integral in providing clarity and understanding of the operational workflows within the HIDC system, emphasizing the structured and systematic interaction between different components to support health data management and analysis.

4.5. Deployment Diagram:

The deployment diagram is crucial for visualizing the configuration of runtime processing nodes and the components that reside within them. This diagram acts as a blueprint for understanding the physical deployment of software components in the Health Informatics and Data Collaboration (HIDC) system.

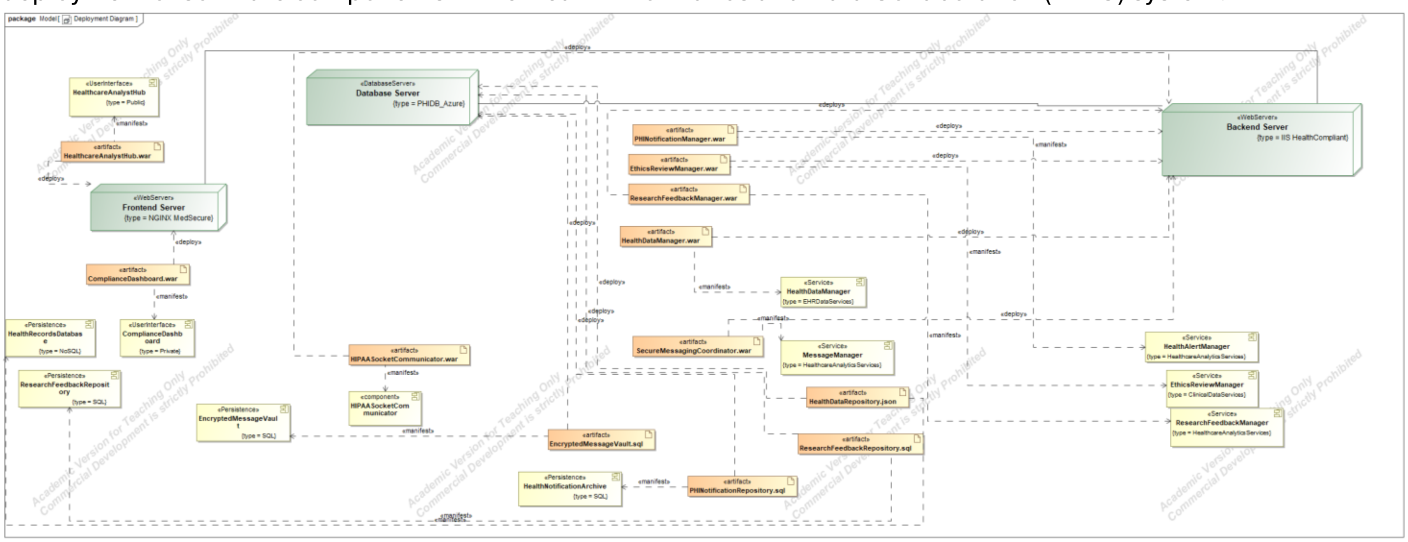

Fig 10: Deployment Diagram for the HIDC system

4.5.1. DEPLOYMENT NODES DESCRIPTION

1. **UserInterface:**

    - The UserInterface nodes represent the user-facing aspect of the HIDC system. Artifacts like HealthcareAnalystHub.war and ComplianceDashboard.war are web applications deployed to web servers, facilitating user interaction with the platform.

2. **WebServers:**

- The diagram highlights a two-tier web infrastructure with Frontend and Backend Server nodes. The Frontend Server handles initial user requests, while the Backend Server manages more complex data processing tasks, maintaining a clear separation of concerns.

3. **Persistence:**

- Multiple persistence mechanisms are indicated, accommodating diverse data types and structures. This includes SQL and NoSQL databases like **HealthRecordsDatabase** and **ResearchFeedbackRepository**, utilizing a hybrid approach for optimal data management and retrieval.

4. **Services and Components:**

- Artifacts such as **HealthDataManager.war**, **MessageManager**, and **HealthAlertManager** represent the various services and components deployed to manage the business logic of the HIDC system. These components are crucial for managing health data, messaging, and alerts.

5. **Communication Pathways:**

- The deployment diagram also illustrates the communication pathways between different components and services, emphasizing their interoperability and the essential data exchange that supports the system's operations.

4.5.2. ARTIFACTS AND MANIFESTS

- Each artifact, whether a .war file for web applications or .sql for database schemas, is linked to a manifest detailing the configuration and dependencies required for deployment. These manifests are vital to ensure that each component is correctly deployed and configured within its environment.

4.5.3 DEPLOYMENT SPECIFICATIONS

- The diagram specifies the deployment of each component onto its respective server, reflecting a strategic distribution of the system's workload. For example, components handling sensitive data, like HIPAASocketCommunicator.war, are deployed on servers with enhanced security measures to safeguard data integrity and privacy.

By documenting the deployment configuration as shown, the HIDC system ensures a clear, structured, and scalable architecture that supports its objectives to facilitate health data analysis and collaboration effectively.

5. CONCLUSION

The Health Informatics and Data Collaboration (HIDC) system is designed as an integrated, comprehensive platform that caters to the multifaceted needs of health data analytics and collaborative research. The detailed visualizations provided through various UML diagrams have meticulously outlined the system architecture to ensure robustness, scalability, and user-centric functionality.

The deployment diagram, in particular, sheds light on the practical application and distribution of software components, encapsulating the intricate relationship between user interfaces, services, and data persistence mechanisms. It underscores the HIDC system's commitment to delivering a secure, resilient, and efficient user experience.

The sequence diagrams provide a dynamic representation of the system workflows, portraying the real-time interactions and data flow between the system's components. These interactions are foundational to the system's operational integrity and are pivotal in facilitating the various functionalities such as dataset browsing, professional communication, research contribution, and feedback management.

As with any system of this scale and complexity, the HIDC system faces challenges. However, employing a robust modeling tool like MagicDraw has laid a solid foundation that supports the system's objectives to streamline health data management and enhance collaborative efforts in the field.

In conclusion, the HIDC system stands as a testament to the power of systematic design and planning. Through the diverse set of UML diagrams, a strategic blueprint has been provided that aligns with the overarching goals of health data analysis and collaboration, paving the way for future enhancements and integration within the broader ecosystem of health informatics.