UNIVERSITÀ DEGLI STUDI DI PADOVA

DIPARTIMENTO DI INGEGNERIA DELL'INFORMAZIONE

Foundations of Databases A.Y. 2023-2024

Homework 1 – Requirements Analysis

Master Degree in Computer Engineering
Master Degree in Cybersecurity
Master Degree in ICT for Internet and Multimedia

Deadline: October 27, 2023

Team acronym	RHO	
Last Name	**First Name**	**Student Number**
Campoy Fernandez	Pablo Julian	2100203
De Faveri	Francesco Luigi	2118936
Groenlien	Iver	???
Padoan	Claudio	2090110
Petrucci	Riccardo	2090302
Vezzosi	Giacomo	2104369

Objectives of the System

Research Hospital in Oncology (RHO) is an institute for healthcare and research in the Oncology field. Given the high volume of medical consultations and the elevated number of clinical analyses conducted at the institute, the medical doctors and researchers affiliated with RHO have identified a pressing need for a well-designed information system capable of effectively managing patient data and clinical analysis information. Specifically, the system should help employees of the RHO center in managing anagraphic information of patients, addressing their consensus for obtaining their data in a GDPR[1]-compliant way and efficiently storing all the information for the research pipeline considering not only the experimental results of the research but also the logistic aspects of such.

Interviews

We conducted several interviews to understand the system requirements in detail:

- The doctors: they clarified how they welcome the patients during the first visit and the kind of information they need to register. They examine their health status and document the patients' information, both anagraphical and first-stage medical analysis, e.g., blood pressure and heartbeat.
- The researchers (clinical engineers and data scientists): RHO's researchers described all the steps of the research pipeline. They explained that biological samples arrive at the RHO's research labs after the medical consultations and need to be properly handled, i.e., stored in *ad hoc* refrigerators and boxes, before the clinical study. After the clinical studies, the results of the analyses need to be properly reported and made accessible for statistical purposes carried out by the Data Scientist at RHO's research labs.
- The Information Technology (IT) technicians: the IT team of RHO explained that they are currently managing all the steps of this process using paper tables, and they have no infrastructure ready, except the hardware systems.

Users and Stakeholders of the System

The **users** of the system are:

- Doctors: they can register new patients, modify patients' general information, and report the kind of samples collected.
- Clinical Engineers: they manage the logistics of the samples, define the characteristics of each clinical study performed, and register the results of chemical analysis.
- Data Scientists: who have permission to query the chemical results collected by clinical engineers in order to carry out a statistical analysis of such outcomes.

The **stakeholders** of the system are the Researchers (Clinical Engineers and Data Scientists) of the RHO Institute that specify the kind of information that the doctors register and define the guidelines of the research studies, clinical and statistical, to be conducted.

[1] General Data Protection Regulation.

Natural Language Sentences

The RHO Institute is in need of a complete and user-friendly information management system to handle its oncology research pipeline for the clinical studies in the hospital, which encompasses the critical functions of patient enrollment, clinical data management, and statistical analysis.

As such, the primary objective of the system is to manage and keep updated the data (both personal and clinical) of patients, the logistics of when and where the research samples are obtained and stored, and the eventual treatment proposed. Moreover, the system needs to be able to group the information based on the clinical study defined by the research team, which takes place with clinical engineers and data scientists for clinical and statistical analysis.

All the hospital's research personnel (doctors, clinical engineers, and data scientists) must be registered to the platform, providing their name, surname, e-mail address, phone number, badge number, and role in the hospital; furthermore, both clinical engineers and data scientists must provide the name of their research team.

Once a cancer patient gets admitted to the hospital, a doctor will first visit the patient, making a diagnosis and proposing a treatment. Concurrently with the visit, the patient will be registered in the system. This operation requires the following information to be provided: name, surname, date of birth, sex, fiscal code, height and weight of the patient, date, type of the visit (or list of the dates and types if there have been other medical appointments), the diagnosis, progression of the illness, state of the patient, and finally the proposed treatment. The patient may be visited more than once by the doctor, who, at each visit, will update the patient's information. During one of these medical appointments, the doctor may take blood or tissue samples to be used for research purposes. In addition, it should be possible to see the patient's clinical history, i.e., the list of sample IDs belonging to the patient already stored in the system.

These samples will be given to clinical engineers, who will first assign each of them a unique incrementally generated identification number, the sample ID, according to the GDPR guidelines for managing clinical information. It is important to underline that a singular sample has the potential to be enrolled in multiple clinical studies. The information gathered from these studies can be either pseudonymized, which involves the removal of Personal Identifying Information (PII) such as name, surname, and city of residence, or anonymized, which implies the removal or masking of the data. The choice to pseudonymize or anonymize the data is determined by whether or not RHO's researchers conduct the study.

The research team, composed of clinical engineers and data scientists, will proceed to initiate the clinical study in a collaborative effort. The clinical study is composed of both: the biochemical analysis, which is the responsibility of the clinical engineers and consists of defining the information belonging to the chemical and biological data of the extracted samples; and the statistical analysis, performed by the data scientists on the available data. If a disease is present in the biological sample, it must be enrolled using the standard for International Classification Diseases for Oncology (ICD-O). The characteristics of the biological data that must be registered are the general features of the sample, e.g., the volume of the sample, whether it is a mutated one or not, the types of cells in the sample, and the number of mutations present, the hormone receptor status of individual cells. Moreover, the system must be able to store information about the clinical results, e.g., tumor marker proteins, oncogenes, DNA and RNA segment characteristics, and specific factors related to the objectives of the clinical study.

Once the data of a sample has been registered, the researchers will provide the location where the sample is currently being stored by specifying the ID of the storage container, such as a refrigerator or a vial. If the sample is being used for experimentation, i.e., the sample is undergoing a specific clinical study by a team of researchers, then the sample will be unavailable.

The clinical results must be made available to the data scientists of the RHO institute in their original form. However, it is equally important to ensure that external researchers receive an anonymized version of the same results. This approach ensures that data is shared ethically, with due regard for privacy concerns, while enriching the amount of data available for research in the oncology field.

Filtered Sentences

Biochemical Analysis, chemical analysis conducted by Clinical Engineers:

- Each Biochemical Analysis is defined by the set of analyses conducted by Clinical Engineers.
- Each Biochemical Analysis must be accessible by the Data Scientist for their Statistical Analysis.
- Each Biochemical Analysis reports the chemical results obtained by the Clinical Engineers.

Clinical Engineer, a member of the hospital personnel involved in Clinical Studies; unlike Doctors, she/he does not have direct contact with the Patients as she/he does not perform Medical Appointments:

- Each Clinical Engineer is defined by her/his name, surname, e-mail address, phone number, badge number, role in the hospital, and Research Group.
- Each Clinical Engineer can log in with a badge number and password.
- Each Clinical Engineer is responsible for labeling and storing the Samples.
- Each Clinical Engineer can define one or more Clinical Studies.
- Each Clinical Engineer can belong to one or more Research Teams for the Clinical Studies.
- Each Clinical Engineer must update the biochemical data of the Sample after each Biochemical Analysis of the Samples.

Clinical Study, procedure carried out by the Clinical Engineers and Data Scientist, by which the biochemical data of the Samples are categorized, registered, and Statistical Analysis is performed:

- Each Clinical Study is defined by the particular type of study done, which can be specified with a Clinical Study ID (CSID) and a brief description.
- Each Clinical Study regards a set of Samples and respective Patients and is carried out by a team of Clinical Engineers and Data Scientists.
- Each Clinical Study is created by a Research Group.

Data Scientist, a member of the hospital personnel involved with processing Patients' biochemical results and clinical data for statistical studies; unlike Doctors and Clinical Engineers, she/he does not have direct contact with Patients or perform Biochemical Analysis:

- Each Data Scientist is defined by her/his name, surname, e-mail address, phone number, badge number, role in the hospital, and Research Group.

- Each Data Scientist can log in with a badge number and password.
- Each Data Scientist has access to the system for the biochemical results of the Samples obtained from the Biochemical Analysis and general information of Patients to carry out Statistical Studies.
- Each Data Scientist can belong to one or more Research Teams for Clinical Studies.

Doctor, a member of the personnel of the hospital that has direct contact with the Patients and is responsible for them, performing Medical Appointments; unlike Clinical Engineers, she/he is not involved directly in Clinical Studies:

- Each Doctor is defined by her/his name, surname, e-mail address, phone number, badge number, and role in the hospital.
- Each Doctor can log in with a badge number and password.
- Each Doctor is responsible for registering the Patients' personal and medical information concurrently with the first visit.
- Each Doctor can perform one or more Medical Appointments for each Patient.
- Each Doctor for each Medical Appointment must update the Patient's medical information.

Medical Appointment, the medical practice, which can consist of either simple visit or surgery:

- The type of service defines each Doctor's Medical Appointment, e.g., introductory visit, routine checkup, visit for collecting Samples, surgery, specifiable with a unique Medical Appointment ID (MAID).
- Each Medical Appointment must comport the Patient's clinical data update.
- Each Medical Appointment could result in the collection of a Sample used for a Clinical Study.

Patient, a person who is admitted to the hospital and has given her/his consent to be visited and for her/his personal and clinical data to be processed.

- Each Patient is defined by her/his name, surname, date of birth, sex, fiscal code, height, weight, clinical history, date of the visit (or list of the dates if there have been other Medical Appointments), the diagnosis (enrolled using the standard ICD-O), state of the Patient (sick or healthy), and the proposed treatment if needed.
- Each Patient can be subject to one or more Medical Appointments.
- Each Patient gives consent for the data to be processed.
- Each Patient releases one or more biological Samples.

Research Team, group of Clinical Engineers and Data Scientists that performs a Clinical Study:

- Each Research Team is associated with one Clinical Study.
- Each Research Team comprises at least two researchers, one of which must be a Clinical Engineer and the other is a Data Scientist.

Sample, the blood, tissue, or genome sequence sample collected by the Doctor during a Medical Appointment and object of the Clinical Study by the Clinical Engineer:

- Each Sample is defined by a Unique Incrementally Generated ID (UIGID), its biochemical data, and the location where it is currently stored.
- Each Sample undergoes one or more Clinical Studies by the Clinical Engineer, where its biochemical data are categorized and registered.
- Each Sample's biochemical features are defined by the Research Team depending on the Clinical Study.
- Each Sample must be registered in one refrigerator or box.
- Each Sample must be available or unavailable.

Statistical Analysis, analysis and results collected by the Data Scientist:

- Each Statistical Analysis is defined by the set of Samples on which the operations are conducted.
- Each Statistical Analysis is accessible by the Research Team of the Clinical Study.
- Each Statistical Analysis reports the results obtained by the Data Scientist.

Term Glossary

Term	Description	Synonyms	Connection
Biochemical Analysis	Analysis performed by one or more Clinical Engineers on a Sample for gathering different results to be processed in a Statistical Analysis.	Biochemical/Chemical Experiments	Clinical Engineer, Clinical Study, Sample, Statistical Analysis
Clinical Engineer	Scientist who enrolls, manages the logistics, and chemically processes the samples of the patients.	Clinical scientist, clinical researcher	Clinical Study, Sample
Clinical Study	Biochemical and statistical analysis performed by Clinical Engineers and Data Scientists.		Clinical Engineer, Data Scientist, Research Team, Sample
Data Scientist	Researcher that performs Statistical Analysis on the clinical results produced by the Clinical Engineer in a Clinical Study.	Statistician	Clinical Study, Sample
Doctor	Medical doctor who takes care of treating patients and enrolls a Patient in the system.		Medical Appointment, Patient, Sample

Medical Appointment	Medical service that a Doctor provides to a Patient to enroll her/his information, update general information about her/him, and (if needed) propose a treatment plan, collecting at the same time a biological Sample for further analyses.	Medical consultation, medical visit, medical examination	Doctor, Patient, Sample
Patient	Person who come to the hospital to have their pathology treated by undergoing a medical examination.		Doctor, Medical Appointment, Sample
Research Team	Group of researchers that perform a Clinical Study.	Research Group	Clinical Engineers, Data Scientist, Clinical Study
Sample	A sample of biochemical material, collected by the Doctor during a Medical Appointment and then analyzed in Clinical Studies.		Clinical Engineer, Clinical Study, Doctor, Medical Appointment, Patient
Statistical Analysis	Analysis performed on the results of Biochemical Analysis by Data Scientists in the context of a Clinical Study.	Numerical Experiments	Biochemical Analysis, Data Scientist

Table 2: Term Glossary table.

Functional Requirements

The database must store:

- **Biochemical Analysis** data including:
  - Operation performed by Clinical Engineers.
  - List of samples used.
  - Results.
- **Clinical Engineers** data, including:
  - Name and surname.
  - E-mail.
  - Phone number.
  - ID (badge number).

  - Its role in the hospital.
  - Its research group name.

- **Clinical Study** data, including:
  - CSID.
  - Brief description of the study.
  - Research team that defined the clinical study.

- **Data Scientists** data, including:
  - Name and surname.
  - E-mail.
  - Phone number.
  - ID (badge number).
  - Its role in the hospital.
  - Its research group name.

- **Doctors** data, including:
  - Name and surname.
  - E-mail.
  - Phone number.
  - ID (badge number).
  - Its role in the hospital.

- **Medical Appointment** data, including:
  - Type of service provided (MAID).
  - If samples were collected or not.
  - The doctor performed the medical examination.
  - The patient that undergoes the medical appointment.

- **Patients** data, including:
  - Name and surname.
  - Date of birth.
  - Sex.
  - Fiscal code.
  - Height.
  - Weight.

  - Clinical history.
  - Date of the visits.
  - Diagnosis.
  - Progression of the illness.
  - State of the illness.
  - State of the patient.
  - Proposed treatment.
- **Research Team** data, including:
  - Name of the group.
  - Clinical studies of the team.
  - Clinical Engineers that belong to the group.
  - Data Scientists that belong to the group.
- **Samples** data, including:
  - UIGID.
  - Biochemical data.
  - Storing location.
  - Availability status.
- **Statistical Analysis** data, including:
  - Methods used by Data Scientists.
  - Data used.
  - Results.

The system must allow:

- Doctors, Clinical Engineers, and Data Scientists to log in and sign up, managing different user roles for them.
- Doctors to register and update patients' personal and medical information.
- Doctor to perform medical appointments for each patient.
- Clinical Engineers to label and store samples.
- Clinical Engineers perform clinical studies for each sample and update their biochemical data.
- Data Scientist to access the biochemical and clinical data of the samples, updating statistical studies.
- Researchers (Clinical Engineers and Data Scientists) to create/join a Research Team.

Non Functional Requirements

The designed system has to:

- Guarantee a quick response to the users' queries.
- Guarantee the persistence of the data through frequent and periodic backups.
- Depending on the Clinical Study, the privacy of the patients must be preserved.
- Provide an easy-to-use User Interface where it is managed the login/registration of the user and the characteristics of the sample, with the possibility to register a new one.
- Support up to 500 insertions of samples per day.
- Support frequent and simultaneous accesses.

Constraints

The system application for this database should satisfy some constraints, namely:

- Implementation using PostgreSQL.
- Client-side implementation as a Web application, using HTML, CSS, JavaScript, and jQuery to ensure usability and maintenance.
- Server-side implementation using JSP and REST web services, Java Servlet, and Tomcat for developing an application in the Java ecosystem.

Group Members Contribution

- **Campoy Fernandez Pablo Julian:** Users & Stakeholders + Term Glossary
- **De Faveri Francesco Luigi:** Objectives + Interviews
- **Groenlien Iver:** No participation.
- **Padoan Claudio:** Functional Requirements + Non functional Requirements
- **Petrucci Riccardo:** NL Sentences + Filtered Sentences
- **Vezzosi Giacomo:** Term Glossary + Constraints