
Nuno Miguel Rodrigues Cantão

# VETIDB – DEVELOPMENT OF AN INTEGRATED DATABASE FOR VETERINARY MEDICINE

Dissertação de Mestrado em Engenharia Biomédica, apresentada à Universidade de Coimbra.

Setembro 2016

· U C ·

UNIVERSIDADE DE COIMBRA

FCTUC FACULDADE DE CIÊNCIAS
E TECNOLOGIA
UNIVERSIDADE DE COIMBRA

Nuno Miguel Rodrigues Cantão

# VetIDB – Development of an Integrated Database for Veterinary Medicine

*Dissertação apresentada à Universidade de Coimbra para cumprimento dos requisitos necessários à obtenção do grau de Mestre em Engenharia Biomédica*

Orientador(es):
Prof. Dr. Ricardo Faustino (La**B**$^3$, Simatix Lisboa)
Prof. Dr. Nuno Ferreira (IBILI/ICNAS)
Dr. Daniel Simões (Clínica Veterinária dos Olivais)

Coimbra, 2016

Este trabalho foi desenvolvido em colaboração com:

# Acknowledgments

Ao Prof. Dr. Ricardo Faustino e ao Prof. Dr. Nuno Ferreira, agradeço a orientação fenomenal prestada durante todo o percurso árduo ao longo do ano, que permitiu que neste momento, haja de facto uma tese de mestrado para defender.

À minha colega de guerra, Sofia Pires, agradeço as muitas horas de companhia e ajuda quando a pesquisa era muita, o progresso lento e as dúvidas e a falta de bases imensas! Juntos conseguimos chegar ao objetivo a que nos propusemos no início deste projeto!

À Ana Nunes e a todos na Simatix e na Clínica Veterinária dos Olivais, particularmente ao Dr. Daniel Simões pela ajuda, sugestões e apoio que prestaram ao longo do desenvolvimento deste projeto.

Aos meus pais, irmão e tia por todas as lições de vida, carinho, apoio e força nos momentos mais difíceis, sem os quais muito provavelmente não teria chegado onde cheguei.

Ao Hugo e à Carolina, agradeço todos os fins-de-semana na terrinha, todos os cafés, saídas para as quais me arrastaram e amizades que tenho a certeza levarei comigo, independentemente do rumo que as nossas vidas levarem.

Ao Henrique, ao Francisco, à Tânia e ao Bruno... Um grande obrigado por todas as risadas, jantares, cinemas, distrações, procrastinação, sessões de jogos, e tudo o resto. Vocês são o grosso do que posso dizer que foi o melhor que Coimbra me deu!

À Filipa, Joana e Blue, amigos, companheiros de casa e confidentes destes últimos anos em Coimbra! Puxaram por mim, tornaram-se no meu lar longe de casa, levantaram-me da cadeira e tiraram-me de casa quando muitas vezes não queria (estou a falar para ti, Blue!

A ti, por acreditares sempre em mim, por todo o amor, apoio incondicional e muita força ao longo dos últimos dois anos, durante todos os bons e os maus momentos.

# Table of Contents

# Abstract

Animal-based research transcends species boundaries. Animal studies are continuously being undertaken, both as a means to gather information to better understand human pathophysiology and in the field of Animal Care, where research is performed with the end goal of studying the animals themselves. As more and more animal data is being generated, it becomes imperative to find ways to effectively store and easily access and analyze this information. While a few human biochemical and imaging databases are available, no such framework exists for animal data. We present an integrated database, VetIDB, which gathers animal data collected during veterinary medicine appointments – namely, digital x-ray images and blood test results of canine and feline patients. This database makes the clinical history of the animal patients readily available through an online platform, to be accessed by both veterinary physicians as a tool for monitoring their patients' evolution, and researchers as a catalogue to choose animal subjects from. By using VetIDB to locate animal-based data of interest, researchers can rapidly cross biochemical information and biomedical images, thus obtaining accurate and reproducible data for both Veterinary Research and animal studies in the field of Human Medicine.

**Keywords:** clinical history, medical imaging, biochemical data, relational databases, veterinary research, PHP, data storage.

# Resumo

A investigação animal transcende barreiras entre espécies. Estudos envolvendo animais estão continuamente a ser realizados, tanto como um meio de recolha de informação de forma a atingir uma melhor compreensão da fisiopatologia humana, como no campo dos cuidados animais, onde a investigação é realizada com o objetivo do estudo do próprio animal. Ao ritmo a que dados de animais vão sendo gerados, torna-se necessário encontrar maneiras eficazes para armazenar estes dados, permitindo que estes sejam acedidos e analisados de forma simples. Enquanto que é possível encontrar alguns sistemas de bases de dados bioquímicos e imagiológicos humanos, o mesmo não se aplica ao caso de animais. Apresentamos então uma base de dados integrada, VetIDB, que tem como objetivo a recolha de dados de animais durante consultas de rotina nos centros veterinários – nomeadamente exames de raio-x digital e resultados de análises ao sangue e urina de pacientes caninos e felinos. Esta base de dados possibilita o acesso ao histórico clínico dos pacientes animais através de uma plataforma online, tanto a médicos veterinários, como forma de monitorização da evolução dos seus pacientes, como a investigadores, funcionando como banco de dados de onde é possível escolher participantes que se adequem às necessidades dos seus estudos. Ao utilizarem a VetIDB para a localização de dados de animais de interesse, os investigadores conseguem rapidamente cruzar informação imagiológica e bioquímica, obtendo assim dados fiáveis e reproduzíveis tanto para investigação veterinária como para estudos animais em campos da Medicina Humana.

**Palavras-Chave:** Historial Clínico, Imagem Médica, Dados Bioquímicos, Bases de Dados Relacionais, Investigação Veterinária, PHP, Armazenamento de Dados.

# Symbols and Abbreviations

CDN – Content Delivery Network

CR – Computed Radiography

CT – Computed Tomography

CVO – Clínica Veterinária dos Olivais (Olivais's Veterinary Clinic)

DB – Database

DICOM – Digital Imaging and Communications in Medicine

DV – Dorsoventral

DBMS – Database Management System

ER model – Entity-Relationship model

FCR scanner – Fujifilm Computed Radiography Scanner

GPS – Geographical Positioning System

GUI – Graphical User Interface

IP – Imaging Plate

MVC – Model View Controller

PACS – Picture Archiving and Communication Systems

RDBMS – Relational Database Management System

SQL – Structured Query Language

VD - Ventrodorsal

# List of Figures

# List of Tables

*"Between animal and human medicine there is no dividing line—nor should there be. The object is different but the experience obtained constitutes the basis of all medicine."*

*Rudolf Virchow (1821–1902)*

# Chapter

# 1 | Introduction

## 1.1 Motivation

### 1.1.1 Pet Ownership Numbers

We welcome pets into our homes for a variety of reasons, including companionship, recreation, and protection[1]. Most pet owners actually refer for their animal companions as being part of the family, being actively integrated with everyday living, participating in family activities and in hard times might even offer support coping, resilience and recovery.[2][3] According to the 2012 U.S Pet Ownership & Demographics Sourcebook, 36.5% of American households own at least one dog and 30.4% at least one cat, spending an average of 378$ and 191$ respectively in veterinary expenses per year.[4]

The benefits of owning a pet are widely documented and acknowledged by many, including the US National Institutes of Health. As a practical example, it is known that one in every three Americans suffer from high blood pressure[5] and according to the American Heart Association, the presence of a pet has a significant and positive effect on their owner's cardiovascular reactivity to stress.[6] It is also worth noting that dogs also appear to positively influence their owners by leading to higher engagement in physical activity and walking, being more likely to achieve the recommended level of physical activity than non-owners.[6] Similarly, there are also arguments in favor of the risks pets can present to their owners. Pets might infect people with disease, cause injuries and even challenge the financial resource prioritization within the owning family.[7]

### 1.1.2 The One Health Approach

We can consider Rudolf Virchow's statement as wise today as it was over a century ago. The idea of absence of a dividing line between the *Homo Sapiens* and remaining species is greatly beneficiary for the *One Health Initiative* defended my many. This initiative consists of a worldwide strategy for expanding interdisciplinary collaborations and communications in all aspects of health care for humans, animals and the environment. It is an effort to properly stimulate the acceleration of biomedical research discoveries and enhance public health efficiency by aiming to improve medical education and clinical care through the expeditious expansion of the scientific knowledge base.[8]

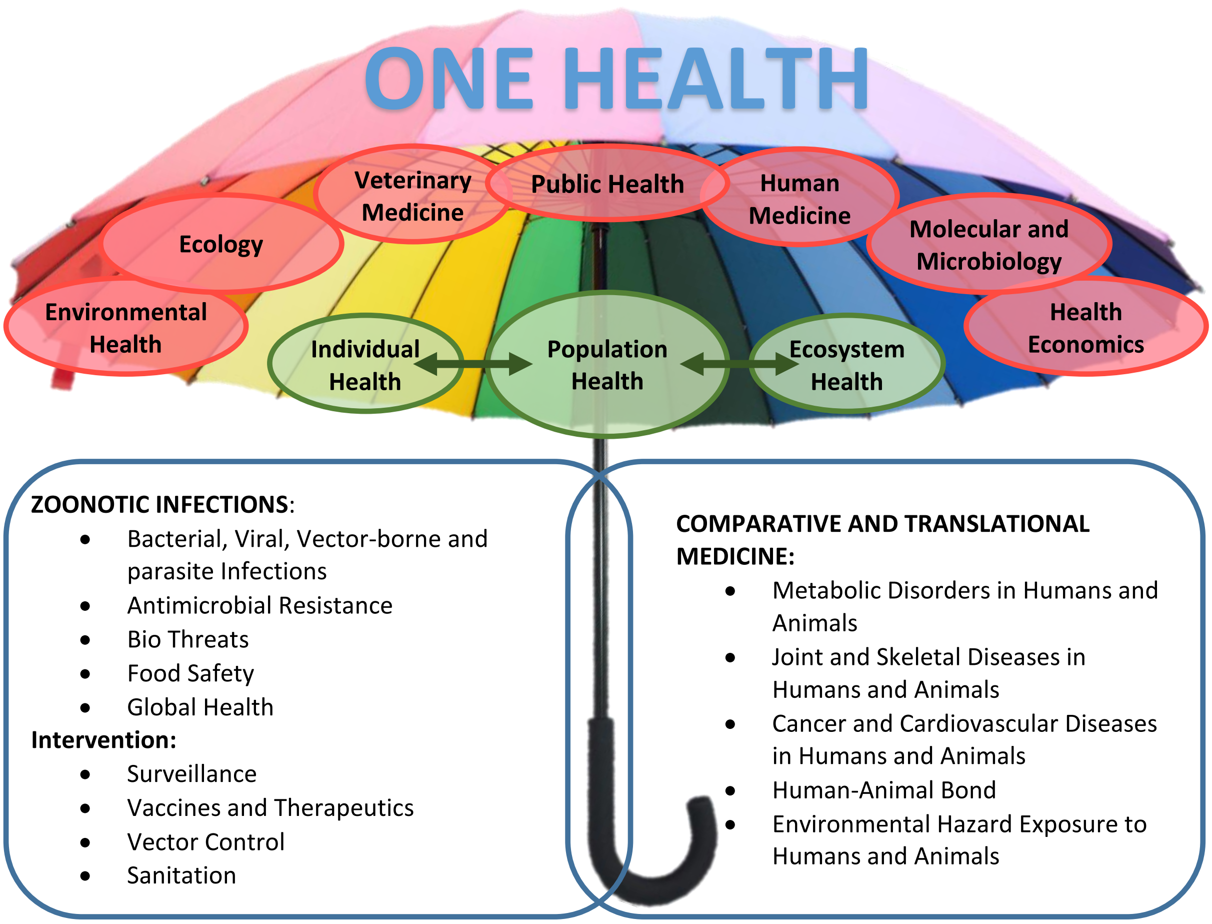



*Figure 1 - One Health Initiative: Interconnections between different study areas affecting not only Human but also Animal health, making contributions to a better state of Public well-being. Adapted from [8]*

The commonalities between different species can and should be exploited since it is a known fact that few diseases affect exclusively one group of animals (wildlife, domestic animals or humans).[9] Zoonotic infections (or *Zoonoses*) are any animal infectious diseases that can be naturally transmitted to humans. These can be caused by a variety of disease pathogens such as viruses, bacteria, parasites and fungi.[10] In fact, of 1415 species of infectious organisms known to be pathogenic to humans, 868 (an impressively high 61%) are zoonotic. Also zoonotic are 132 out of 175 (75%) pathogens associated with emerging diseases[1].[11] Based on these premises, one could argue that veterinary medicine and research are fundamentally human health activities.

### 1.1.3 Contributions of Veterinary Research

The historical contributions of veterinary research to advancements in medicine are rich and widely documented. From the early days of medicine where comparative anatomical and physiological studies were conducted in order to provide a basis for our understanding of embryonic development, human blood circulation and lymphatics, organ structure and function to Pasteur's experiments on rabies and anthrax vaccination in sheep or Koch's studies of tuberculosis (all of which depended either on the knowledge of comparative or veterinary research).[9] However, it has been in the last two decades that the vital role of veterinary research has been brought into stark reality. Concerns driven by the recognition that many emerging infectious diseases of humans are zoonotic (such as *Escherichia coli* infection, bovine spongiform encephalopathy, avian influenza, Ebola virus disease, salmonellosis and influenza[11]) highlight the importance of research in order to improve veterinary public health and food safety.

### 1.1.4 Data Storage Needs

Several gigabytes of medical information are generated on a daily basis. A substantial part of this information corresponds to imaging data: the Geneve University Hospital alone reported the production of 12000 medical images per day [12] and, in 2006, around 377 million Radiology examinations and 18 million Nuclear Medicine procedures were performed solely in the United States. [13] As increasingly larger amounts of medical images (as well as other

[1] Diseases or pathogens are considered as emerging when they appear in a human population for the first time, or have occurred previously but are increasing in incidence or expanding into areas where they had not previously been reported.[11]

patient data) are being produced, having an efficient and economic data storage system is crucial.

While some efforts to increase the efficiency of human clinical data storage systems are being made, the same – surprisingly – isn't true for animal data. Animal studies are continuously being undertaken, not only in Veterinary Medicine, but also as an important pre-clinical step in human pathophysiology research.

In Veterinary practice (at least, where Portuguese veterinary centers are concerned), it is common to have separate storage solutions for each individual type of patient information. For instance, a veterinary clinic's x-ray imaging records can be stored in a specific digital archive, while the patient's clinical history is kept in a different (and incompatible) computerized archive system and their blood test results are filed away on yet another format. As a consequence, valuable animal patient data – suitable for studies in both Veterinary Medicine and Human Health Research – are not being properly stored, integrated and systematized and are incredibly vulnerable to system failure and data loss.

## 1.2 Objectives

With this project we are aiming to create an Integrated (relational) database that acts as a cloud-based archive for animal imaging and biochemical information in order to address the needs of a proper storage solution in veterinary medicine.

This database aims not only to offer an all-purpose archive and organization tool for the different types of data generated during the Veterinary practice, but also to provide a complete online platform for Veterinary research. By combining the animals' clinical history and patient demographics with their imaging and biochemical data, VetIDB constitutes a unified and complete database for researchers to select their data from. In addition, it will facilitate data sharing amongst different Veterinary centers and practitioners, thus contributing to a better cooperation between different Veterinary teams, so that an improved and personalized patient care can be offered to animals and pet owners.

In order to meet the expectations of veterinary practitioners, a partnership was established with CVO in order to assess information collected in regular consults and the standard veterinary practices. Data from the clinic's patients that had both imaging and biochemical data was also collected in order to populate and test VetIDB.

## 1.3 Structure of this work

This work is comprised of 7 chapters and is organized as follows.

| Chapter | Brief Description |
|---|---|
| **1 \| Introduction** | Through this first chapter, a brief introduction stating the motivation as well as main objectives of the work conducted throughout the year are presented. A short summary of the contents of this work is also made. |
| **2 \| Theoretical Background** | This chapter focuses on the theoretical background of the work conducted through the last year, presenting a brief introduction to medical imaging, databases and DBMS and also databases in the medical practice. The fundamentals about Relational Databases and it's conception are also discussed. |
| **3 \| Equipment Used and Data Collected** | Chapter 3 focuses on the data collection process at the veterinary clinic that hosted this project. Some of the equipment and limitations of the technologies used at the facilities are explored. |
| **4 \| Database Design & Implementation** | The fundamentals and process of conception and implementation of a relational database are explored as well as the steps that go from the design of the |

| | |
|---|---|
| | Conceptual Model to the actual final database. |
| **5 \| Choice of Technologies** | The process and justification of the software choices made during the development of this work are explored in this chapter. Both for the RDBMS and the development framework for the application. |
| **6 \| The Platform** | In this chapter, a presentation and functionality description of the final work developed throughout the year is made. Topics such as user authentication, layout presentation and functionality as well as the process of adding and accessing data are explored in this chapter. |
| **7 \| Conclusions/Future Work** | This chapter marks the end of the work. A brief conclusion about the work conducted throughout the year as well as a small discussion of the weaknesses and future potential of the platform are also explored. |

*Table 1 - Summary of contents explored throught this work.*

# Chapter

# 2 | Theoretical Background

## 2.1 Medical Imaging

The definition of Medical Imaging is generally accepted as the techniques and processes that allow the creation of visual representations of the interior of a body for clinical analysis and medical intervention. It seeks to establish a database of what is considered to be normal anatomy and physiology with the aim of allowing the identification of possible deviations or abnormalities. Depending on the context it can be considered as a sub-discipline of biomedical engineering, medical physics or medicine. The use of diagnostic imaging has increased significantly over the last two decades. The upsurge of new and improved technologies and techniques of medical imaging can be credited as a main reason to the earlier and more accurate diagnoses of disease via non-invasive methods. Payments to physicians for diagnostic imaging exams in the US have the highest rate of growth amongst all of the physician provided services. However, CT and nuclear medicine examinations use a higher dose of ionizing radiation than conventional radiographs and the rapid growth of the use of some of these imaging techniques has led to an estimation (supported by extensive epidemiological evidence) that 2% of future cancers will result from current imaging use, if imaging continues at current rates.[14][15]

Although a lot can be said about medical imaging and its contributions to diagnosis and research, considering the scope of this project, we will be focusing mainly on Computed Radiography.

### 2.1.1 Computed Radiography

The history of X-rays dates back to 1895. Wilhelm Conrad Röntgen, a Professor at the Wuerzburg University in Germany, discovered the X-ray and noticed that, while able to pass

*Figure 2 - First radiograph taken. In it, we can see the hand of Roentgen's wife. As he discovered, the unknown radiation at the time was able to pass through human tissue but unable to go through his wife's bone structure or metal wedding ring.*[16]

through human tissue, it was unable to pass through bone or metal. The name “X” was actually given in order to identify it as an unknown type of radiation[14]. This discovery would grant him the very first Nobel Prize in Physics[17]. Radiography is, therefore, an imaging technique wherein electromagnetic radiation other than visible light (especially X-rays) are used in order to observe the internal structure of non-transparent objects of varying density and composition. A good example of this is the human body. In the simplest of terms, an image is obtained when an heterogeneous beam of X-rays produced by a generator is projected towards the object which ends up absorbing a part of it. The remaining X-rays that are able to make it through are captured behind the object using a detector (usually either photographic film or a digital detector). This originates an image that is a superimposed 2D representation of the object’s internal structures.

CR is a type of X-Ray imaging. However, instead of using traditional photographic film, it uses an imaging plate (IP) made of photostimulable phosphor housed in a special cassette and placed under the body part to be examined. Then, by using a special laser scanner (CR reader), the image is read and digitized. This is the imaging technique most used at CVO for diagnostic purposes as it allows for an almost instantaneous development of images on a digital display allowing for the digital transfer and enhancement of images (applying filters, zoom, adjusting contrast, brightness).

CR also usually requires up to 30% less dose than film. This can be explained by the fact that CR usually requires fewer retakes due to under or over exposure, resulting in a lower overall dose to the patient.[18] It also allows for image previews to be generated in less than 10 seconds. The adjustment of image brightness and/or contrast allows for a wider range of thicknesses to be examined with a single exposure. The same cannot be said about conventional film based radiography, where a different exposure or multiple film speeds in one exposure are required in order to cover wide thickness range in a component.

*Figure 4 - Feline Lateral Thorax view taken using CR at CVO*

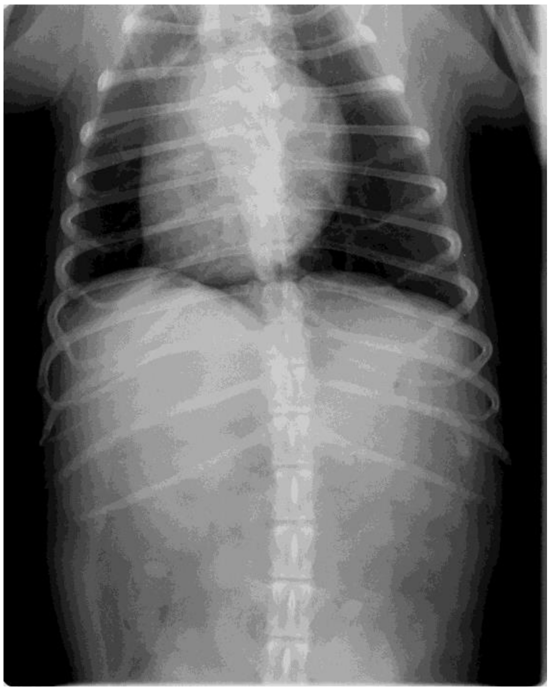


*Figure 3 - Canine Ventrodorsal (VD) or Dorsoventral (DV) view taken at CVO using CR (in small dog and cat thoracic radiographies the differences beween DV and VD are not well visualized[19]*

## 2.2 Databases

The current definition found in the Merriam-Webster Dictionary tells us that a database is “a usually large collection of data organized especially for rapid search and retrieval (as by a computer)”[20]. Formally, a database consists not only on a collection of related data but also to the way it is organized. For any major business, data is an asset and the efficiency in how we input, retrieve and use such data directly affects our ability to faster achieve either success or failure. A proper database in fact alleviates the ancient need for a human to search through millions of filing cabinets for a single record. They have become the standardized and performance friendly way to store, retrieve and analyze data. Commodities we have taken for granted such as the Internet, GPS, Electronic Banking and much more, simply wouldn’t be possible without the existence of databases.

The evolution of technology and tremendous advancements in the areas of computer processors, memory, storage and networks have led to improvements of massive magnitude in the capabilities, sizes and performance of databases and their respective database management systems (or DBMS for short).

## 2.3 DBMS

A DBMS is a software package that allows the creation, definition and manipulation of databases, while also providing tools to perform any kind of operation on the data on them. Some of the basic functions that DBMSs offer their users include[21][22]:

1. Persistent storage of large sets of data;
2. Creation and schema definition of new databases (logical data structure);
3. A command-line or graphical user interface that allows the user to access and/or modify the data. This can be achieved either with tools of the user interface or SQL language. SQL (Structured Query Language) is a special-purpose programming language designed specifically for the management of data held in relational DBMSs It can also be used for database schema creation and modification;[23][24]
4. A management system for transactions, responsible for multiple simultaneous access by several different processes. In database context a single logical operation on the data is called a transaction. However, a transaction might bundle several different operations. A good example of this is a transfer between two bank accounts. The whole action is considered a single transaction, however, several operations/steps occur (including debiting one account and crediting the other). A true transaction must adhere to the ACID properties in order to offer guarantees that avoid the end user of much of the headache of concurrent access to mutable database state.[25][26] ACID is an acronym that stands for:

    a. <u>A</u>tomicity – The concept that each transaction must be "all or nothing": if any of the singular parts of a transaction fail, then the entire transaction fails. The database state remains the same. In order to be considered an atomic system, atomicity needs to be guaranteed regardless of the situation, including power failures, crashes and errors.

b. Consistency – Ensures that all transactions will lead the database from a valid state to another. It validates any data written to the database according to defined rules (constraints, data types, cascades, etc).

c. Isolation – Property that guarantees that the concurrent execution of transactions results in a system state that would be exactly the same if the transactions were to be executed serially. With isolation, an incomplete transaction cannot affect another incomplete transaction.

d. Durability – Once committed, a transaction will persist and will not be undone to accommodate conflicts with other operations. In relational databases, for example, when a group of SQL statements execute, the results have to be stored permanently even in the case of an immediate crash. To avert against unpredictable situations (such as a power loss) the transactions and their effects need to be recorded into a non-volatile memory[2]

With technology evolving at such a rapid pace, obviously better and more complete DBMSs have come a long way since entering the market in the late 60's. Many more functionalities have become a staple of modern DBMSs, including redundancy control, profile and data restriction functionalities (allowing for the definition of different roles/users, that are able to access part of the information stored) and allowing simple backup and restore solutions in case of hardware of software failures. Some examples of popular DBMSs are MySQL, PostgreSQL, Microsoft Access and Oracle DB.

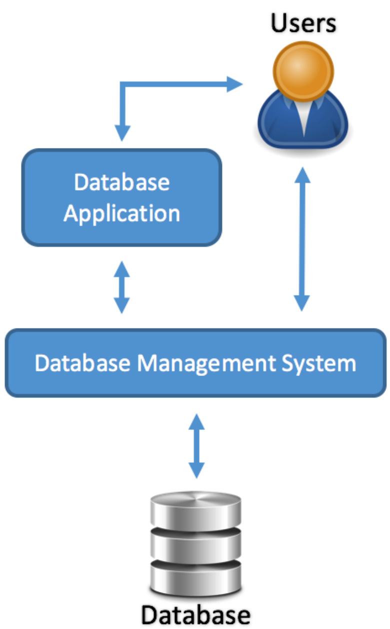



*Figure 5 - Access to data contained in a Database can be done directly using a DBMS or rather through a more personalised Database Application*

[2] Non-volatile memory is the name given to all computer memory that can store information after being power cycled (turned off and back on). Examples of this include magnetic hard-drives, floppy and optical discs. An example of volatile memory is RAM (Random Access Memory) which loses all of its contents once the computer is shut down.

## 2.4 Databases in the Medical Practice

Every year, over 98,000 Americans die due to medication errors alone. In order to address this, and many other issues, the healthcare industry is quickly adopting the use of databases to track everything from the prescription of medicine and laboratory results to patient outcomes.[27] However, medical practice involves innumerous healthcare professionals, patients, protocols and procedures, exams and treatments and each of them has to be registered, described and either has or generates its own information. As stated before, several gigabytes can be generated on a daily basis at a small imaging center alone.

Picture Archiving and Communication Systems (PACS) are the preferred storage method for imaging data in the clinical practice.[28] PACS-like systems allow for the incorporation of both images from different medical imaging modalities and non-image data, such as the patients' clinical information. Whenever a patient undergoes a medical imaging examination – say, an x-ray or a Computed Tomography (CT) scan – their newly generated imaging data can quickly be filed away in the PACS archive of the medical institution.

In the past few decades, the advancements in the field of Bioinformatics have generated new approaches and tools for data storage and organization: remote database services for the storage of human biochemical data[29] and medical images[30] are now available. Although such online archives are particularly useful when dealing with large datasets – as is the case with medical images, which enclose massive amounts of information, encoded by and expressed as numerical values – they are mostly used for research purposes, and never as the default data archiving system in the clinical practice.

Fortunately, Health Care professionals are now becoming more receptive to remote image storage methods,[31] mainly because cloud-based archives stand as an economically feasible solution for the ever-expanding footprint of medical imaging data, as the number and information-density of these images keeps growing over time. Depending on the local server's size and capacity, PACS-like systems entail installation, configuration, maintenance and technical support expenses, not to mention hardware updates every five years (as most local servers start to lose the ability to adapt to increasing workloads after approximately four years) and an average salary of $60,000 for a system administrator. An equivalent cloud storage solution would result in an average monthly saving of 79%.[32]

As an example of this move to cloud based solutions we can talk about CUF. CUF is a vast and growing Portuguese network of private hospitals and clinical practices. It is very common to be exposed to publicity to their hospitals either with television or radio commercials and street or magazine advertisements. These add campaigns actually highlight the company's data archives. All of the patient's consults, exams and clinical history is stored and shared between all of their facilities. This obviously is a tremendous advantage in case a patient needs to be examined by a different doctor or a different facility. A mobile application is also made available to costumers where appointments can be made and exam results and other clinical information can be downloaded which can tremendously improve patient care and treatment, even if the patient seeks treatment somewhere else.

## 2.5 Relational Databases

Relational databases are digital databases organized according to the relational model of data, proposed by E. F. Codd in 1970. They are maintained by special versions of DBMSs called relational DBMSs. Virtually all relational databases use SQL as the language for query building and database maintenance. The idea underneath relational databases is that each table is a set of sets (or rows). Within a single table all of these sets must have the same data structure, containing a list of named fields (or attributes) and their respective values.

| id | Given_name | Family_name | Date_of_birth | Grade |
|---|---|---|---|---|
| **1** | John | Ives | 02/12/1989 | 3.96 |
| **2** | Steve | Jones | 10/11/1988 | 2.76 |
| **…** | … | … | … | … |

*Table 2 - An example of the classic row-column approach to tables*

One might claim that a spreadsheet would achieve the same effect. However, the resemblance is only superficial. Database tables offer much less flexibility than spreadsheets since they require that all the rows within a table have to possess the same structure for their data and calculations generally take place outside of the tables, not within them. From this healthy obsession with neatly organized data comes the power of Databases. Even though some applications can offer a view much like a spreadsheet grid of their tables allowing for a simpler way to edit their contents, in reality, data is being moved in and out of a table using SQL code, always respecting the rules defined in the table structure or *schema* like the one in Table 2.

| Field name | Data type |
|---|---|
| **id** | integer |
| **Given_name** | string |
| **Family_name** | string |
| **Date_of_birth** | date |
| **Grade** | float |

*Table 3 – A Schema specifying data types for each field in Table 1*

Depending on the database further details to the schema can be specified such as size limits, minimum and maximum values and even whether the field can or not be left without a value. Even though some web applications might only require a single table, for more complex solutions, working within a single table is just too constraining.

If we look closely at the data in *Table 1*, there is a field called id that is unique to each row. This is called a primary key. A primary key is a field in a table that uniquely identifies each row in a database table. The primary key has to be unique to each field and therefore can be either sequentially generated or, in the example given, could be a unique identification card or social security number since we know that there is no possibility of having two records with the same identifier. By having unique identifiers within a table, we can establish connections between tables in a simple way: by simply referring to the other table row by including its primary key. A field that identifies a row in another table is called a Foreign Key. An example is shown below.

| id | Given_name | Family_name | Date_of_birth | Grade |
|---|---|---|---|---|
| **1** | John | Ives | 02/12/1989 | 3.96 |
| **2** | Steve | Jones | 10/11/1988 | 2.76 |
| **3** | Sarah | Parker | 12/08/1989 | 4.62 |
| **...** | ... | ... | ... | ... |

| id | Award | Year | Student_id |
|---|---|---|---|
| **147** | Best Smile | 2014 | 1 |
| **159** | Cleanest Desk | 2016 | 3 |
| **197** | Outstanding Generosity | 2016 | 3 |

| id | Presenter | Award_id |
|---|---|---|
| **74** | John Cusack | 147 |
| **89** | Meryl Streep | 197 |
| **100** | Julia Roberts | 159 |

*Figure 6 - An example of how tables might be connected/related in a database. The Student_id and the Award_id fields are used to identify a row in another table, therefore we call these fields Foreign Keys.*

This way of connecting tables avoids the insertion and storage of redundant data in the database. If we wanted, for example, to store all of the awards and students information in a single table, each student's information would have to be inserted as many times as the awards he wins. These links also allow us to make simple questions such as: "Which Presenters gave awards to John Ives in 2014?" and obtain the answer "John Cusack". A program is easily able to follow the Id's and links between them to come up with answers to this type of questions.

This example represents a very simple way of connecting tables within a database. More complex types of links can be established and sometimes even extra tables need to be created in order to represent links between two different tables, by storing the primary keys of each table in a new table as foreign keys.

### 2.5.1 Development of a Relational Database

The development of a Relational Database involves design, implementation and testing stages. It is a very complex and time consuming task that usually follows the process detailed (very briefly) below.

As with every software application, the first step is the requirements elicitation. These requirements should be as detailed and thorough as possible in order to assess the needs of the architecture of the database. This is followed by the creation of a conceptual data model reflecting the structure of the information to be held. This should be a concise description of the data requisites that the end users have made and include the type of entities, relationships and restrictions. This can be achieved by developing an entity-relationship model, often with the help of drawing tools/software (one example of this is the PowerDesigner Software suite). It is in this step that all errors and conflicts should be ironed out in order to guarantee the integrity of the relational diagram. Deep questions about the things of interest should be asked in this state, for example: "can a client have more than one animal?", or "can a veterinary practitioner also be a client/animal owner?". The answers to these questions will define the entities/tables, their attributes and the relationships between them.

After choosing an appropriate DBMS, the conceptual model is transformed into a logical model or schema. The logical data model, unlike the conceptual data model is dependent on the DBMS chosen. During this stage, a methodical approach called normalization should be applied. It's goal it to ensure that no information is recorded in more than one place, in order to maintain consistency after any insertions, updates or deletions.

At the end, decisions affecting performance, scalability, recovery and such should be made. This is often referred to as physical database design. Physical data modeling is highly depended on the current state of technology and is most likely going to be quite outdated in few years' time. By distinguishing between the conceptual and physical models, we can separate the more stable from the unstable parts of a design.[21][22][33]

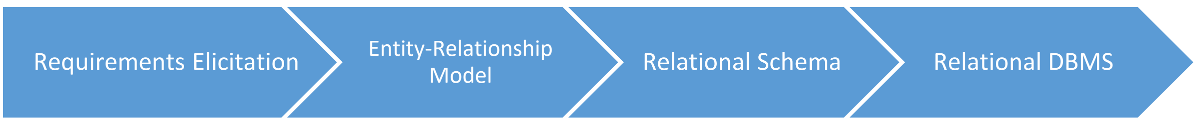



*Figure 7 - The Process of Conception of a DB. Adapted from [22]*

# Chapter

# 3 | Equipment Used and Data Collected

## 3.1 Data Selected to Populate Database

By partnering with CVO, it was possible to better understand the Veterinary Medicine practices and usual consult routine. Since the aim was to create a database able to store both imaging and biochemical data, data was collected from dogs and cats that had been submitted to both computed radiographies and blood sample analyses. A brief description of the equipment used to acquire both types of exams can be found in the next section.

### 3.1.1 Imaging Data

The Computed Radiography equipment used at the clinic is the FCR Capsula X made by Fujifilm®. As described in the theoretical background section, images are acquired using imaging plates (called FCR cassettes) that are then digitized used the FCR Capsula X and are able to be viewed and digitally enhanced by using the bundled software in the workstation (FCRView).

*Figure 8 - CR Cassette being introduced into the FCR Scanner in order to be digitized to the Workstation.*

Even though the function was present in the software, DICOM files were unable to be extracted from the workstation. The only possibility to extract and store the imaging files was in jpeg file format. Attempts were made in order to contact the Fujifilm® support staff but no answer was obtained regarding this situation. However, VetIDB was designed with the DICOM standard in mind and DICOM files can be uploaded and stored in the database and a DICOM viewer is included in the online platform and will be further discussed in the next chapter.

### 3.1.2 Biochemical Data

The Mindray® BC-2800Vet is a compact and fully automated hematology analyzer used at CVO. The results of the blood sample analysis are printed for each patient through the use of a small integrated thermal printer. All of the results are then filed away in a folder and kept at the veterinary clinic. All of the data from the patient's that had also performed imaging exams was organized, typed into digital format and then used to populate the database. Some of the parameters obtained included GLU-PS, TP-PS, ALB-BPS, GPT-PS, GOT-PS, AMCYL-PS, BUN-PS, CRE-PS, Lymph#, Mon#, Gran#, Lymph%, Mon%, Gran% and Eos%.

### 3.1.3 Additional Patient Information

Patient information is stored using a software suite called QVET. QVET is an integral management software system for veterinary clinics[34]. It manages not only patient's medical records but also the veterinary practice's stocks, products and services. It is a very capable, functional and complete solution for veterinary practices. However, what we noticed during our time at the clinical practice is that information is not always completely registered at the animals' clinical history. Since the imaging and biochemical data each have their own storage and organization methods, information is dispersed and isn't always easy to relate and find. During the data collection process, several identification numbers didn't match between QVET, the FCR workstation and with the hematology reports. Some imaging and biochemical data couldn't even be associated to a patient in QVET (and therefore pertinent information such as the patient's age, weight and diagnosis couldn't be obtained). Some records had to be manually completed by matching the dates between imaging exams and consults. Due to a hard drive failure, some patient's records were also missing, which turned the data collection into an inefficient and time consuming matchmaking task.

# Chapter

# 4 | Database Design & Implementation

## 4.1 Schema Creation and Methodology

In order to define and implement VetIDB, several steps were followed, each of which will be detailed in a section of this chapter:

1. Requirement Elicitation, aiming to define the functionalities and needs the platform should give a solution to.
2. Definition of the Database Structure using a Conceptual Data Model consisting of an ER model. During the duration of the project, this diagram suffered several alterations, aiming to better reflect and answer the needs of the solution, always taking into account the logical associations between all the data handled.
3. The choice of a RDBMS was made taking into account the type of data to be stored and the limitations of each type of DBMS. Several options were considered and the winning choice will be detailed and justified through the course of this chapter.
4. A Physical Data Model was generated and used to create the Database Schema.

It is important to know that VetIDB didn't always set out to be the consult report tool that it ended up becoming. The first idea for VetIDB was simply an offline database to store and easily access both imaging and biochemical data of animals in order to facilitate the discovery of new biomarkers. After observing the consult routine, the amount of dispersed information and different storage systems at the clinic, we realized that time is a precious resource at a veterinary practice. In order to give veterinary practitioners the proper motivation to introduce an extra step in their current routine we had to give them something back. Hence the idea of turning VetIDB into an online consult report tool. Veterinary practitioners can use the platform as a constantly updatable and up-tp-date clinical history of their patients with all of their information stored in a unified and single solution and, at the same time, researchers would have access to imaging and biochemical data and chose subjects that best fit into their research projects.

## 4.2 Requirement Elicitation

Aiming to give a proper answer to the objectives stated at the beginning of the project, the Database should comply with the following elicitations:

1. The data organization must be able to effortlessly link several different types and number of exams to a single patient and ensure that every exam belongs to one, and only one patient.
2. Adding new fields to the main entities of the database should be possible and a relatively simple process. Medical information is mutable. Relevant data can be found everywhere and the database needs to be able to adapt and evolve to mirror this concept.
3. Different types of users should have different types of access to the information. This serves not only to protect patient information but also to avoid the removal or editing of information by someone that does not have the privilege to do so
4. Although adding information will be the main used function, editing and removal of information contained in the database must also be possible.
5. Searches allowing users of the database to easily find patients or exams by age, sex, breed, diagnostic or report commentary must be possible.

## 4.3 Use Cases

In this section, the use cases diagram for our platform following the UML (Unified Modeling Language) standard will be presented. These restrictions are implemented in the online platform that is used to access the database.

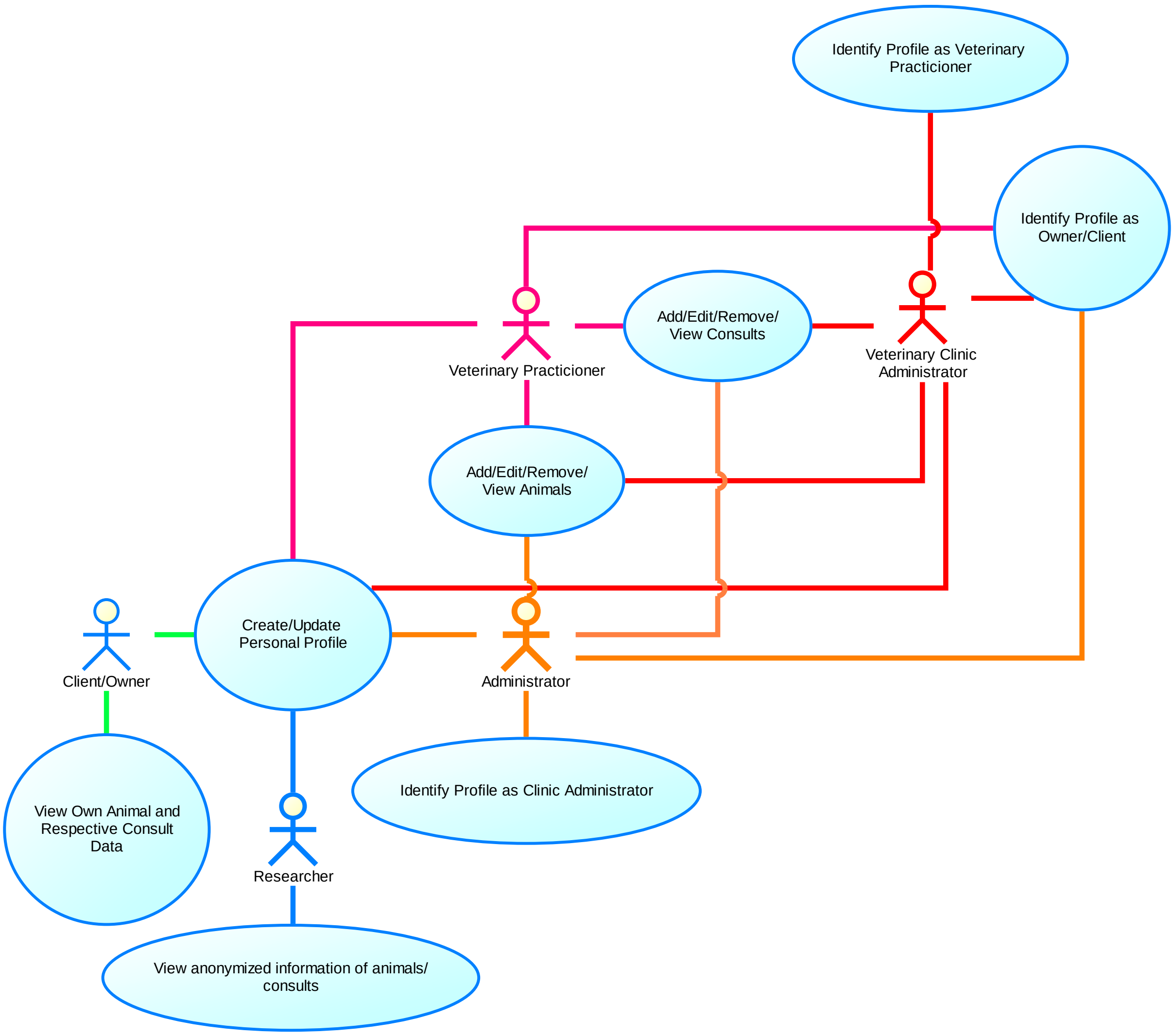



*Figure 9 - Use Cases Diagram for the DB. (Diagram Designed using the Sybase PowerDesign 15)*

As can be seen in the Figure 9, 5 different actors (or roles) were defined. The roles of administrator, Veterinary Clinic Administrator, Veterinary Practitioner, Client/Owner and Researcher. All of them can create and update their personal profile while other functionalities are highly dependent on the role they are assigned. For example, a Researcher will be able to view anonymized data from patients included in the database, while the client will be able to fully view all information pertaining the animals associated to him. None of these two roles are authorized to make any alterations to the information regarding the animal's information.

A veterinary practitioner, a veterinary clinic administrator and the system administrator will all be able to identify a profile as a client/owner and associate the animals that belong to him. The system administrator will be able to identify a profile as a clinic administrator for a veterinary practice and both of them will be able to set profiles as veterinary practitioners. Veterinary practitioners, and both types of administrators will also be able to add/edit/remove animal/consult data.

## 4.4 Conceptual Model

The development of a conceptual model was done through the use of the Sybase PowerDesigner software. An ER model was established but due to its size and complexity the figure corresponding to the database schema is only included in the Appendix section of this work.

We have a total of 84 fields stored in 13 different entities and the design process behind the choices made regarding the data organization and links between tables will be explained throughout this section.

Each user of the Database will have an entry in the profiles table where all of the personal data and information will be stored. This is also where login information (e-mail and password) will be saved. Being an online application, passwords shouldn't be stored in the database in plain text. A process called hashing is done to the user's password, which means that what is actually stored in the password field in the profiles table is the result of a one-way function to encrypt the user's password. There is no "reverse" function to reverse or undo a hash, so it isn't possible to simply use the hashed password and try to obtain the unencrypted password.

We then have the entities Clients, Administrators, Clinic Administrators and Vets. Each profile can be associated to one (or more) roles. This association is made by entering the profile id into the respective table(s). There's a "one-to-one" type of connection between the table profiles and each of these different entities, meaning that each Vet/Client/Clinic Administrator/Administrator must have one and only one profile. The reverse isn't exactly the same since a profile might belong to a vet but it might also represent a client. This is the case where the Vet at one Clinic might be the Client at another clinic (for example).

There are other rules of association between the Administrators table and the Clinic Administrators and Clinics. For example, Several Clinics can be run by several clinic administrators. This will have repercussions when the physical data model is generated since a new table will be created due to the type of connection in order to store the clinic administrators responsible for which clinics. Clinics also have a relationship with vets. They are employers of the clinics and will store the primary key of the practice they work in as a foreign key in the vets table. A practice can have one or more vets, but a vet must always be associated to one and only one clinic.

Between the animals and clients table, there is a “many-to-many” relationship, meaning that an animal can belong to one or more clients and a client can have one or more animals. This will also lead to a new table being generated in order to store the “owning” information, associating animals to their owners.

Animals might need consults, therefore a “one-to-many” relationship links these two tables. An animal might need one or more consults, but a consult can only be linked to a single animal. It is in the consults table that the information subject to change over time is store. The animal’s weight, for example is something that doesn’t make sense to store in the animal’s table since it is constantly changing thorough the course of the patient’s life.

The same “one-to-many” relationship is done between consults and each of the different type of exam. A consult might require one or more radiography exams or hemograms. But each exam has to be associated to one and only one consult.

A radiography exam also has a link to the incidences table. During the time spent at the clinic, we noticed that during a radiography session, more than one incidences were taken (for example, a canine lateral thorax and canine lateral abdomen incidences could be performed during the same radiography session). Therefore, Between the Radiographies and Incidences table there is a “One-to-many” relationship. Meaning, a radiography can generate one or more incidences, but an incidence must belong to one and only one radiography.

It is also important to note that, in order to better optimize the storage of imaging files, these shouldn’t be stored directly in the database. Rather, the file’s path in the file system should be stored in the database, and it should be loaded when necessary.

## 4.5 Physical Data Model

The physical data model corresponds to a logical model where restrictions in the naming conventions and implementation are already taken into account considering the choice of RDBMS. It is, however, a very similar scheme to the conceptual model. We can see that a few extra tables needed to store the information concerning the relationships between certain tables are added (for example between the animals and clients tables).

The physical data model was once again generated using Sybase PowerDesigner, it is done automatically but, in the process, looks for inconsistencies and errors in the conceptual model design. The Physical Data Model generated can be viewed, once again, in the appendixes section of this work.

## 4.6 From the Physical Data Model to the actual Database

After successfully obtaining the physical data model it is time to transform it into SQL code that will create the tables in the RDBMS system and apply the restrictions that will guarantee data integrity for the platform.

The Sybase PowerDesigner software will also be able to generate the appropriate SQL code to our choice for RDMBS. In the case of VetIDB, we chose to use PostgreSQL. It is a very powerful cross-platform, yet open-source RDBMS that perfectly suits our needs for the platform envisioned.

After running the SQL code in PgAdmin III (PostgreSQL's graphical user interface administration tool) we can perform all sorts of operations on the database (Figure 10). It is important to note, however, that all use cases and restrictions are yet to be applied since that will be managed by the online platform that we developed to access the database.

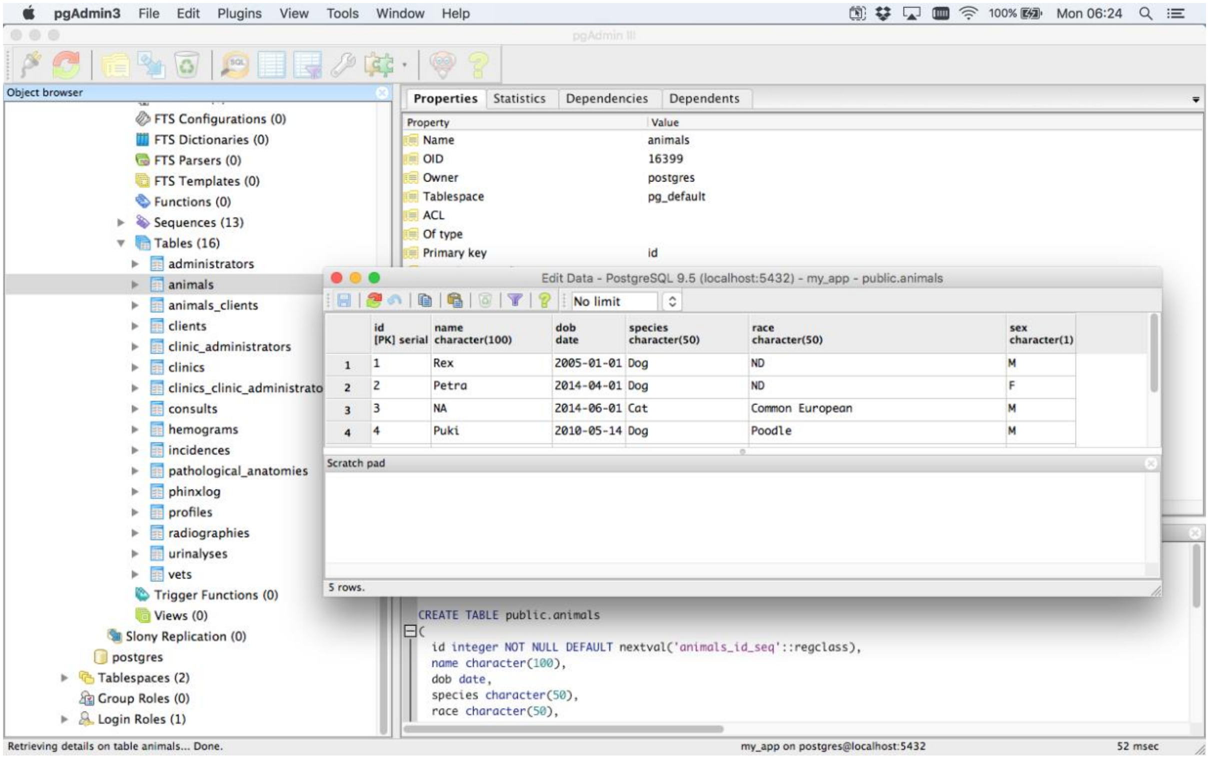


*Figure 10 - Using PGAdmin III in order to access the Database. All of the tables are created and some data has been used in order to populate it. In this window we can view a small extract of the animal's table.*

# Chapter

# 5 | Choice of Technologies

## 5.1 PostgreSQL as our RDBMS

PostgreSQL is a powerful, multi-platform and open source relational database system. It has been in active development for over 15 years and it has a proven architecture known for its reliability, data integrity and correctness.[35] Some of its appealing features include the unlimited database size and enormous size per table (32TB of data). In our application we will not be storing imaging data directly in the database, but as stated previously, a lot of data is generated due to medical practice. The fact that it is an open-source solution also helps to narrow licensing costs down and since it is multiplatform and can run virtually in any host, expenses with hosting can be quite lower (specifically when compared, for example, to Microsoft's SQL server, that only runs on windows machines, which implies extra expenses on the server side). It is also ACID compliant in all of its transactions. The importance of this was explained in the Theoretical Background (Chapter 2).

While there are other popular (and even open-source) options in the market such as MySQL, for example, it also comes down to a little bit of personal taste. Both have their own advantages and limitations. For example, MySQL won't validate the size of a VARCHAR field and will often just truncate it (eliminate part of the data). PostgreSQL, on the other hand, will not allow data to be inserted if it is too long. MySQL used to be regarded as the fastest between the two and PostgreSQL as the richest in functionalities and reliability. However, recent comparisons have stated that both companies have made tremendous improvements and are, in fact, approaching each other both in features and in speed.

## 5.2 CakePHP as a Development Framework

CakePHP is an open-source web development framework that follows the MVC approach and is written in PHP. It follows well-known software engineering concepts and software design patterns such as convention over configuration and association data mapping.[35]

MVC is a software architectural pattern originally used in the development of user interfaces on computers. Major programming languages have adapted the MVC as an architecture for Web Applications.

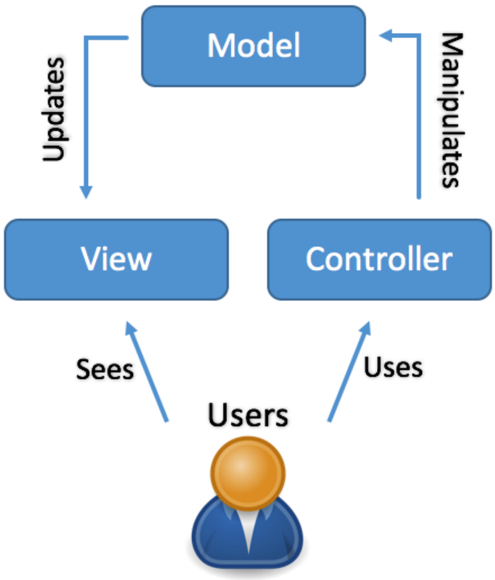


*Figure 11 - Typical Collaboration of the MVC Components. Adapted from [36]*

As can be understood by its name, the MVC is build up by three main components:

1. Model – The model is the part of the application responsible for managing the data of the application. It represents the business logic of the application and is responsible for retrieving and converting data. It is responsible for processing, validating, associating and other tasks related to the handling of data.
2. View – Views render a presentation of the data in a particular format triggered by the controller's decision to present the data, in our specific case, the View is an HTML page.
3. Controller – Part of the application responsible for handling user input/requests and perform the interactions with the data model objects. It receives an input, validates it and then performs the operation that ends up modifying the state of the data model.

CakePHP follows a very strict set of naming conventions (reflected in the database's entities and field names) that make the beggining process of building a web application more of a preparation than coding job (hence the use of convention over configuration). By following the conventions set by CakePHP, we can use their scaffolding tools which will set up the very basic components needed to access and manipulate data from a given Database.

The conventions set by CakePHP for database naming are the following[37]:

1. Table class names are plural and CamelCased. Animals, ClinicalAdministrators, PathologicalAnatomies are all examples of conventional model names;
2. English words are to be used in the naming of table and column names, this is used in order for CakePHP to process the right inflections (from singular to plural and vice-versa) CakePHP also supports the adition of other languages rules for some words by using a special utility case called Inflector, however that wasn't exactly necessary for the purposes of our application since English works just fine;
3. Field names with two or more words are underscored: first_name, last_name.
4. Foreign keys in hasMAny, belongsTo or hasOne relationships between tables are recognized as the name of the related table, followed by id. So Animals hasMany Consults, the Consults table will refer to the Animals table via a Animals_id foreign key;
5. Join tables, used in BelongsToMany relationships between models should be named after the model tables they will join (Animals_Clients);

## 5.3 CakePHP setup and Database Connection

Setting up a new CakePHP project is a relatively simple task as long as all the requirements are met. These include a web server and a copy of CakePHP's folder structure.[38] The simplest method of obtaining the latest version is through a tool for PHP dependency management called Composer.[39] This project will make use of several pre-existing plugins to assist in functions such as searches and file upload/management. Composer makes integrating these libraries easier and assures that when it is time to deploy the project to a server, all dependencies are met.

For development and testing purposes, an Apache Server with PHP v.5.6.23 was installed on a personal computer. Both the database and online platform are hosted in the same machine. This is by no means a limitation on the final version. CakePHP allows for a very smooth and simple setup to remote database hosts and will not require much effort when it is time to deploy the application into a remote server location.

CakePHP supports the following relational database servers:

1. MySQL 5.1+
2. SQLite 3
3. PostgreSQL 8+
4. SQLServer 2008+
5. Oracle (through a community plugin)

As mentioned in the previous chapter, VetIDB was creating using PostgreSQL as its RDBMS. Connecting to the (in this case local) database is done by editing the values in the Datasources.default array in the config/app.php file with the ones that match our setup. These include the username and password to access the RDBMS, host, database name and the driver matching the RDBMS we're using.

By following the conventions set by CakePHP, we can have a (simplistic) version of the platform up and running in very little time. This is done using the bake console. The bake console will generate the controllers, models, views that match our database's tables. After this step we have a very basic, albeit functional application that provides data access to our database.

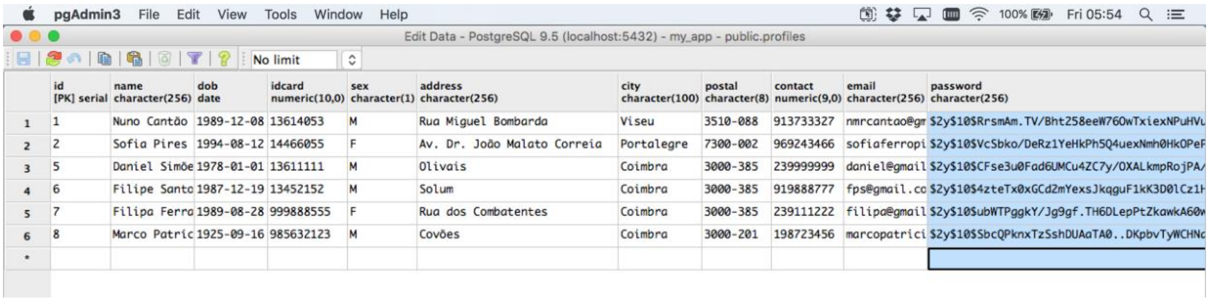
# Chapter

# 6 | The Platform

## 6.1 User Authentication

Authentication in CakePHP is handled via the AuthComponent. It allows for the combination of authentication and authorization objects in order to create flexible ways to handle user authorization.

In our application, authentication is done using e-mail and password as credentials. Passwords are not, however, stored as plain text in the database as that would imply serious security flaws. Instead, the AuthComponent provides tools for password hashing. When a Profile is created, the password chosen is hashed and then stored in the database.




| | id [PK] serial | name character(256) | dob date | idcard numeric(10,0) | sex character(1) | address character(256) | city character(100) | postal character(8) | contact numeric(9,0) | email character(256) | password character(256) |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | 1 | Nuno Cantão | 1989-12-08 | 13614053 | M | Rua Miguel Bombarda | Viseu | 3510-088 | 913733327 | nmrcantao@gm | $2y$10$RrsmAm.TV/Bht258eeW760wTxiexNPuHVu |
| 2 | 2 | Sofia Pires | 1994-08-12 | 14466055 | F | Av. Dr. João Malato Correia | Portalegre | 7300-002 | 969243466 | sofiaferropi | $2y$10$VcSbko/DeRz1YeHkPh5Q4uexNmh0HkOPeF |
| 3 | 5 | Daniel Simõe | 1978-01-01 | 13611111 | M | Olivais | Coimbra | 3000-385 | 239999999 | daniel@gmail | $2y$10$CFse3u0Fad6UMCu4ZC7y/OXALkmpRojPA/ |
| 4 | 6 | Filipe Santo | 1987-12-19 | 13452152 | M | Solum | Coimbra | 3000-385 | 919888777 | fps@gmail.co | $2y$10$4zteTx0xGCd2mYexsJkqguF1kK3D0lCz1H |
| 5 | 7 | Filipa Ferra | 1989-08-28 | 999888555 | F | Rua dos Combatentes | Coimbra | 3000-385 | 239111222 | filipa@gmail | $2y$10$ubWTPggkY/Jg9gf.TH6DLepPtZkawkA60w |
| 6 | 8 | Marco Patric | 1925-09-16 | 985632123 | M | Covões | Coimbra | 3000-201 | 198723456 | marcopatrici | $2y$10$SbcQPknxTzSshDUAaTA0..DKpbvTyWCHNc |
| * | | | | | | | | | | | |

*Figure 12 - A view of the data in the Profiles Table using PGAdmin 3, PostgreSQL GUI for accessing data in databases. Passwords are hashed before being stored in the database as a security precaution.*

In order to use the information stored in the Profiles table as login credentials, methods for login and logout had to be created in the ProfilesController file. A login screen (or view) was also programmatically created in order to provide the user a friendly screen where authentication could be performed. The information inserted by the user is compared against the credentials stored in the database and upon successful authentication, users are (by default) redirected to their personal profile screen. If, however, the information inserted does not match any profile in the database, an error message is shown indicating the username or password inserted are incorrect (Figure 13).

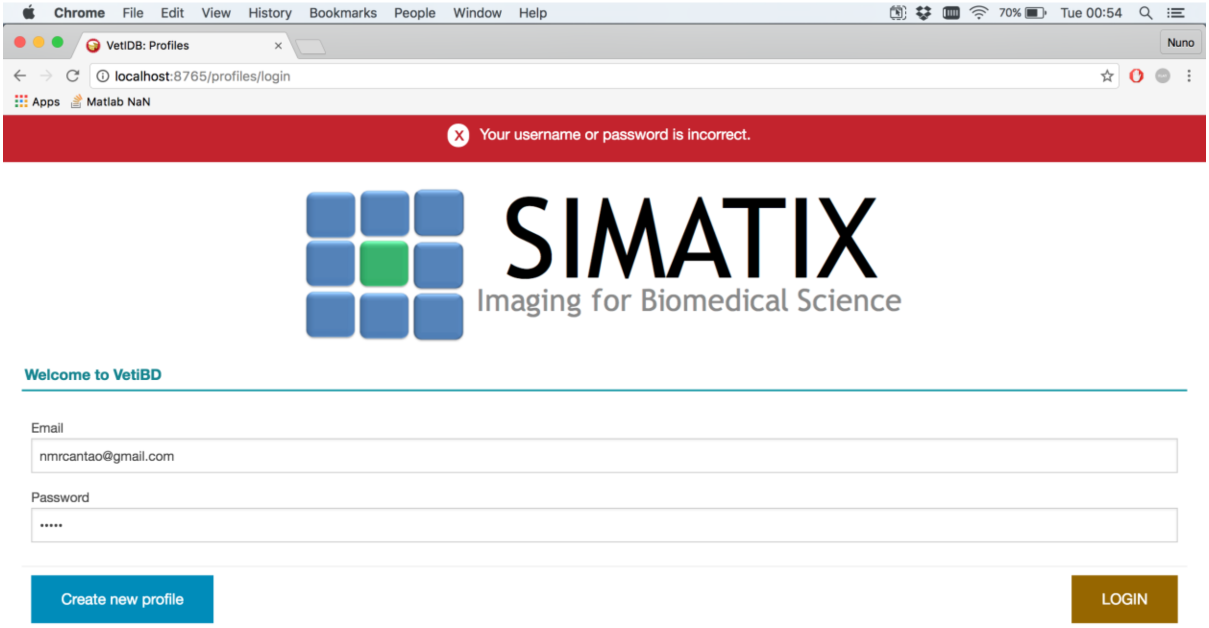



*Figure 13 - Authentication unsuccessfull. Credentials that were inserted by the user didn't match any profile in the database.*

A new profile can also be created from the welcome screen in order to access the database. By default, this profile will not be associated to any of the user types contained in tables of the database (Administrators, Clinic Administrators, Vets, Clients) and will, therefore, be classified simply as a researcher profile.

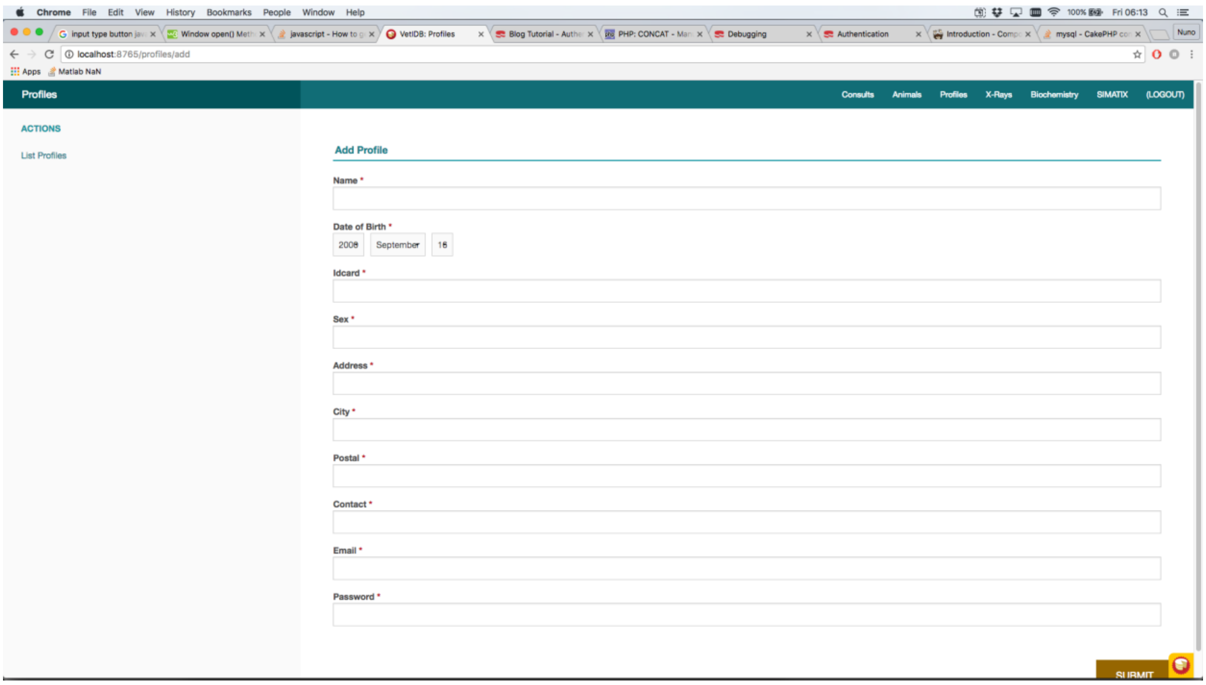



*Figure 14 - Create new Profile screen. By default, accessible by anyone who does not posess credentials to log into VetIDB.*

## 6.2 Layout Presentation and Menu Functionality

After a successful login, users are redirected to their personal profile page. From it, users can access a variety of options accessible through two different menu systems.

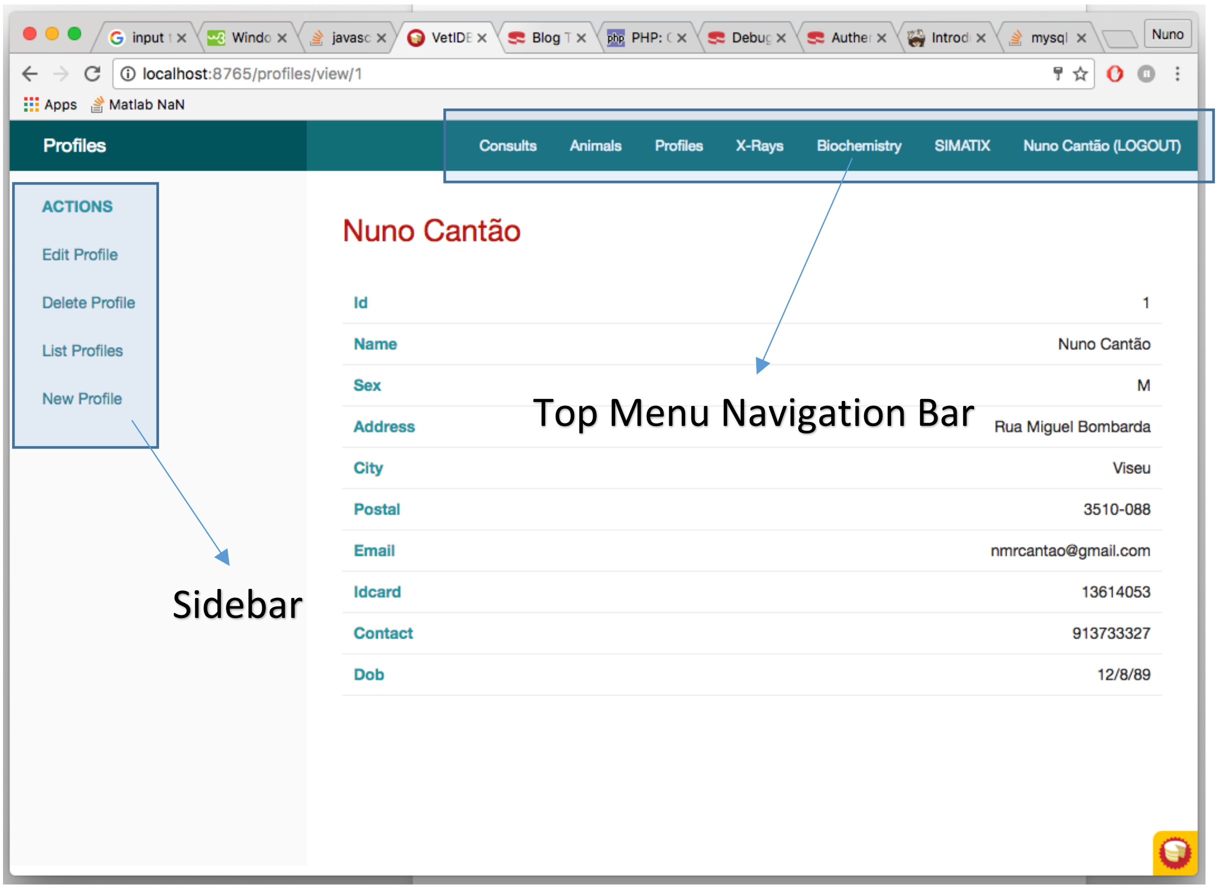



*Figure 15 - The two menu systems present in the application.*

In Figure 13 we can observe both navigation systems implemented in the platform. The top navigation bar allows for quick access to the index pages of each of the main entities in our database. Consults, Animals, Profiles, X-Rays, Biochemistry, a link to the Simatix website and a Logout button that shows the name of the currently logged-in user. This top menu remains the same in every single page of the platform, for every type of user. Time constraints have limited the personalization of every screen functionality for every type of user and, therefore, when a logged-in user tries to access a functionality that isn't allowed for their role, a warning message is displayed in the top bar warning the user that he isn't authorized to access a certain location. The sidebar options change depending on the current section the user is in. Therefore, for example, while browsing the Animal's List, instead of functionality concerning Profiles, the sidebar will display options to add a new animal, add a new client or perform a new consult.

## 6.3 Defining Roles & Data Relationships

Every single user of the database will have a profile. However, profiles can belong to users with different types of roles and privileges. In our Database model we handle this differentiation through the use of different tables for each user role (except for the role of researchers, but that will be explained later). By adding the profile ID as a foreign key to the clients, administrators, clinic administrators or vets tables, we are identifying that profile as one of those roles. That doesn't exactly mean that one profile cannot be identified as more than one role (a clinic administrator might also perform consults at their clinic and, therefore, needs to be identifies as a veterinary practitioner too since the entries in the consults table require the id of the veterinary that performs a consult as a foreign key.

System Administrators are able to create new clinics, and new clinic administrators, creating the adequate associations between them. Clinic Administrators can identify profiles as veterinary practitioners and associate them to their clinics. Both types of administrator and veterinary practitioners are able to identify a profile as a client at a veterinary clinic. As a client can have multiple animals, these relationships can also be established when making the association between a profile and a client.

By establishing these associations between animals, clients, profiles and by accessing the information of one of these elements, the connected information can quickly be obtained without big effort, this simple flow of information can be seen in the next figure (All the data contained in these pictures is mock data. It is made up for the purposes of this work and does not consist of any data collected at CVO, only the medical images used in this theses were collected at the clinic, however, patient's names have been altered and client associations also do not correspond to reality).

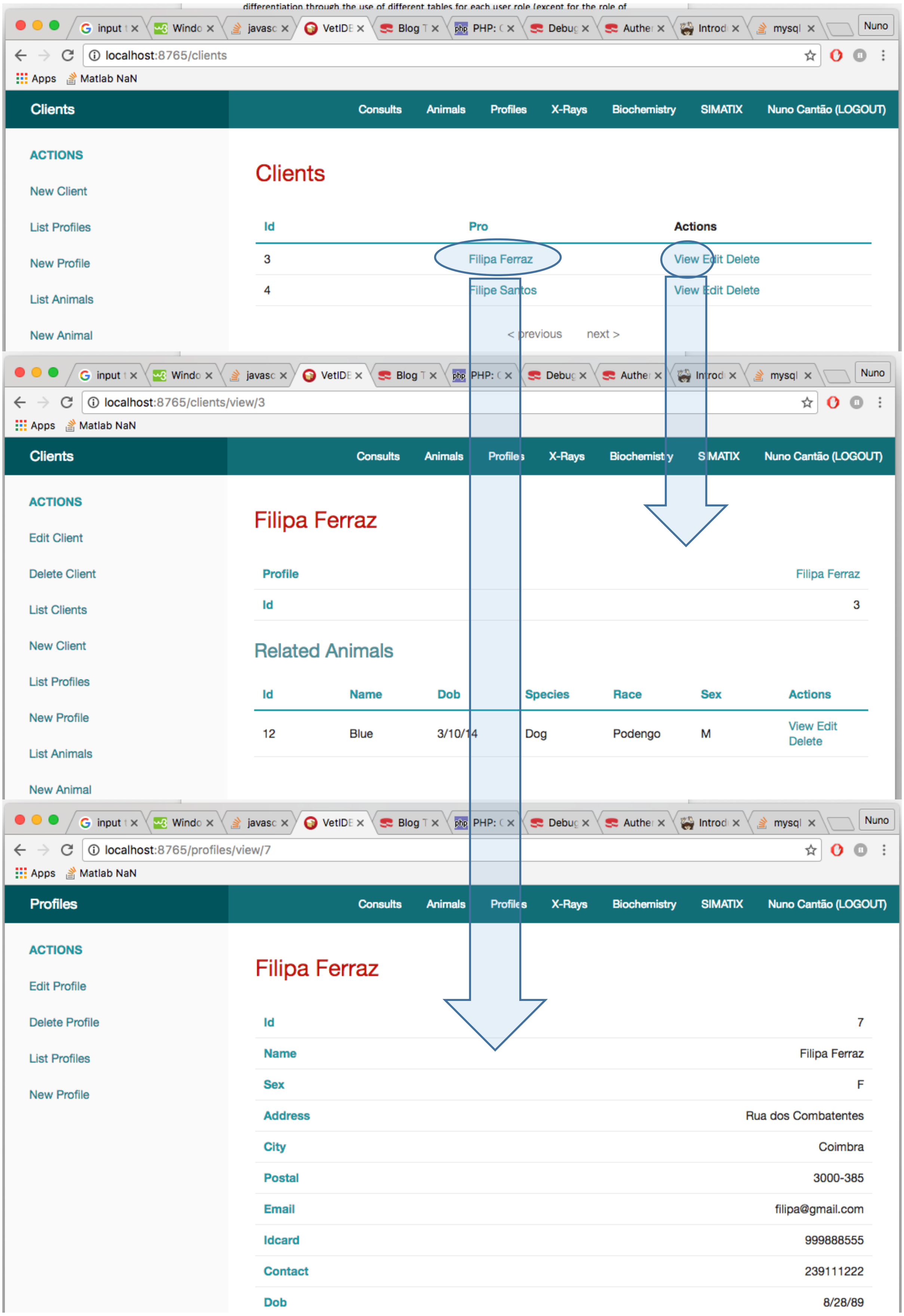



*Figure 16 - The Power of Relational Databases: From a single client screen we can view the related animals to a particular client, the clients profile and even access the animal's patient history by exploring the correct relationships between the different tables.*

## 6.4 Animal Data, Consults and Exams

The main purpose of VetIDB is to provide an online platform where data collected from the veterinary medicine practice can serve as a pool of data for veterinary research. This is why building a reliable and as complete as possible clinical history of each patient was of the utmost importance in the planning stages. By using relational databases, we have proven that related information can easily and quickly be displayed. The same attention was taken into account when dealing with consultation records and exams performed at the veterinary practices.

When a new patient comes into the clinic, an animal profile is created for him, registering all the information that isn't time dependent.

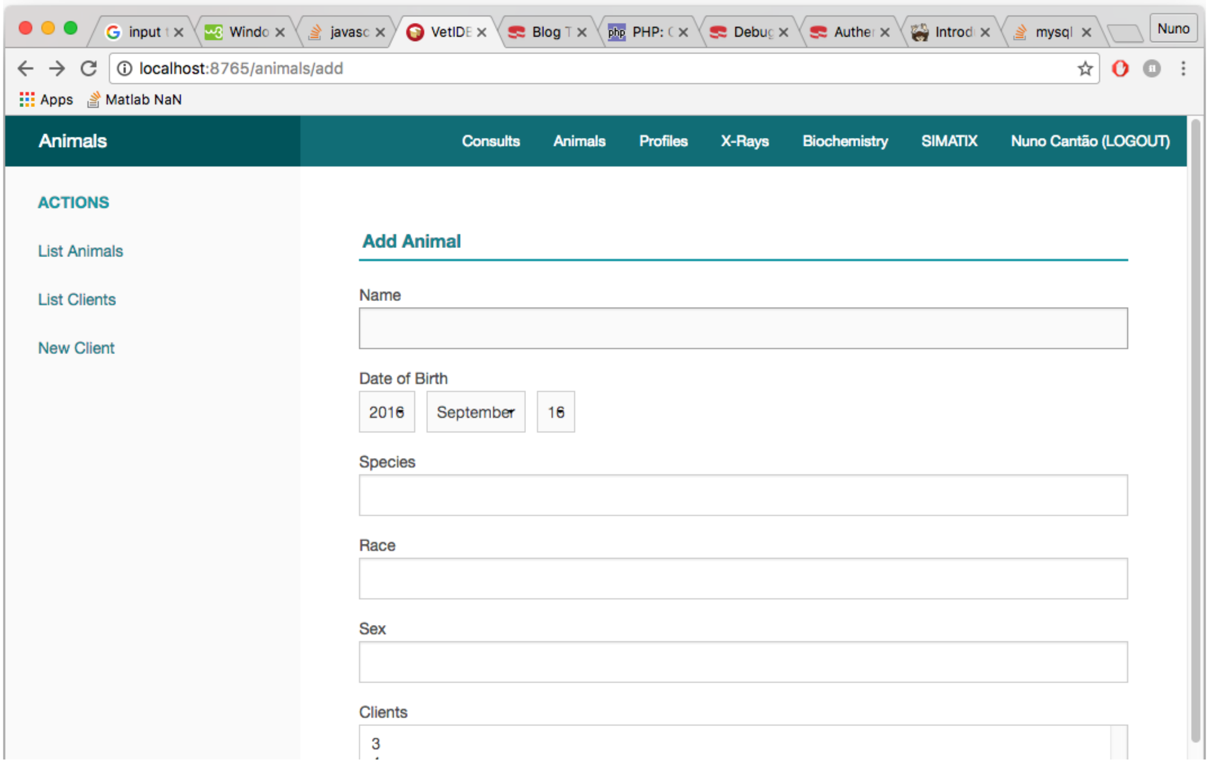



*Figure 17 - Adding a new animal. Time independent information such as the Date of Birth, name, species, race and sex is registered upon arrival at the clinic.*

Associations to client profiles can be made in this step using the client's id or can be made later through the client page. This type of relationship isn't mandatory since there are times when animals are treated without having been brought in by any clients.

With each visit to the clinic, a new consult object is created where variables such as the animal's current weight, heart rate analysis and temperature are measured and registered.


*Figure 18 - Every consult can connect via a single click to the Veterinary Practicioner's information and other consults. It is also possible to quickly access the animal's file and complete history of consults.*

### 6.4.1 Registering and Storing Exams

Just like in human medicine, in veterinary medicine sometimes exams need to be performed in order to facilitate diagnostic. As stated previously, VetIDB aims to be a platform able to store both imaging and biochemical data in order to provide accurate and reproducible data for both Veterinary Research and animal studies in the field of Human Medicine. The way exam results are stored into the database is through the association to a consult. The consult itself is associated to one and only one animal, making it easy to establish links between animals, consults and exams.

Taking for example a CR exam. When a patient comes to the clinic for a consult and a CR exam is performed, a new Radiography exam is associated to the consult id. Each exam might require one or more incidences. The individual patient's clinical history will connect to every single consult he ever attended along with an individual report from each visit to the clinical practice. Each consult will also be associated with all of the different exams performed on that consult. If we look at the last screen on Figure 18, we can observe that there is an option to view the details of Rex's consult (consult id number 9).

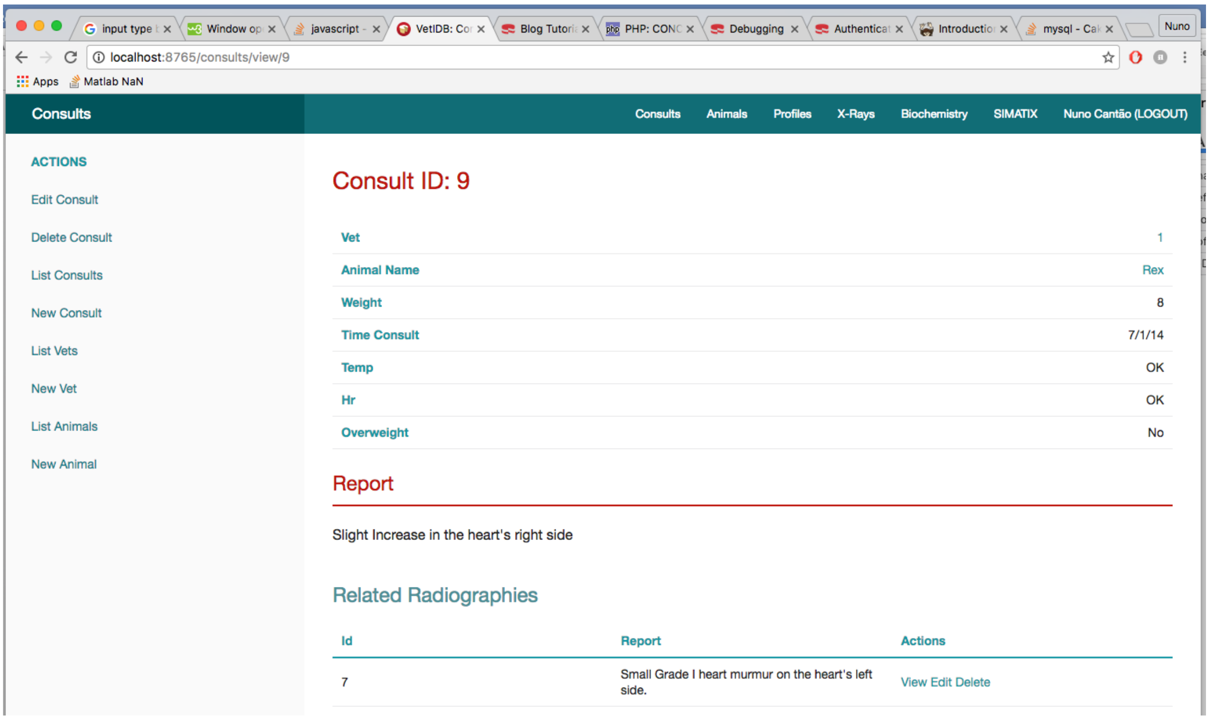



*Figure 19 - Consult with ID 9. A radiography exam was performed during Rex's visit to the clinic.*

Each different image acquired during the radiography exam is called an incidence. A radiography session can include several of these. In Rex's case, however, only one incidence was taken.

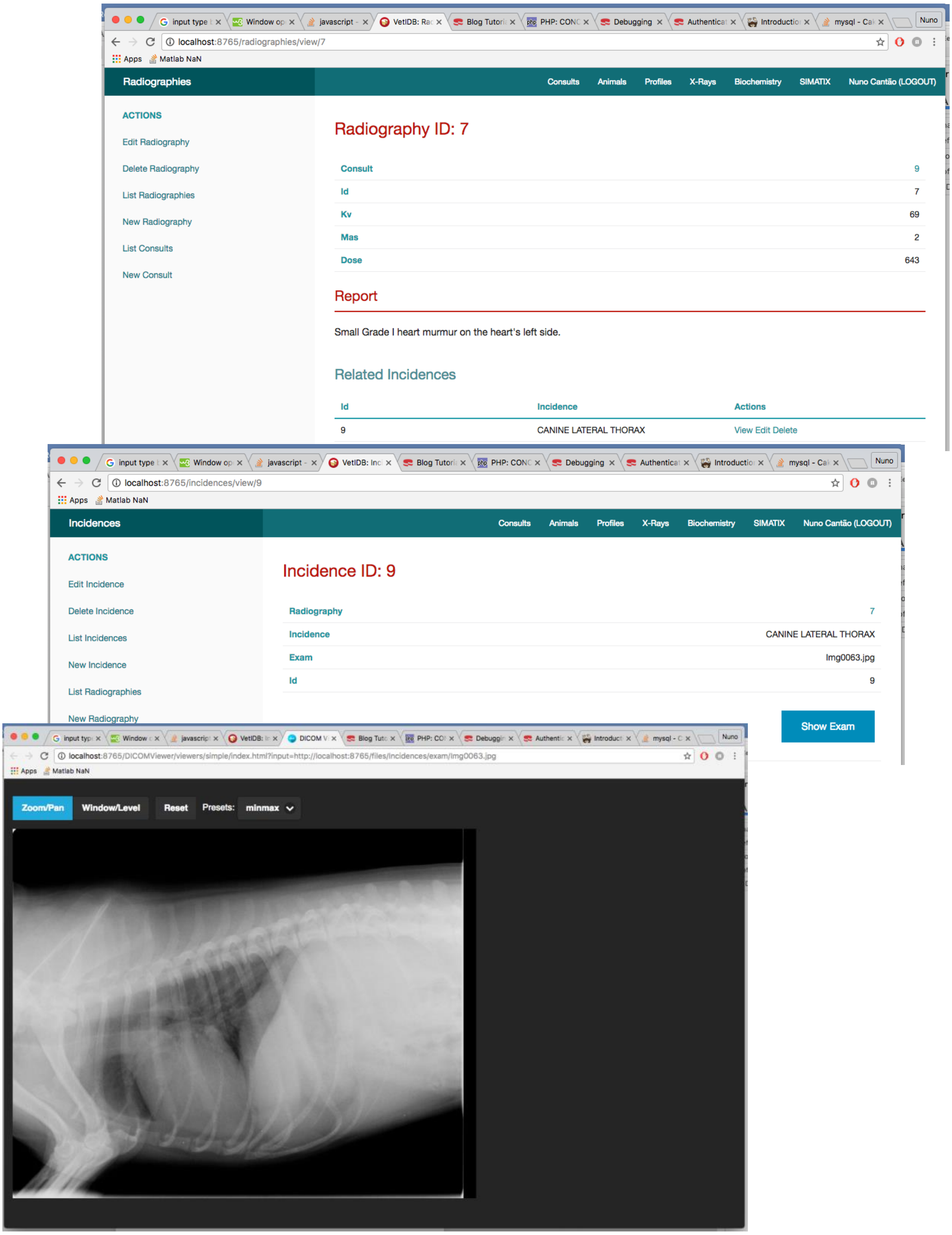



*Figure 20 - Going from the first screen to the last, from the radiography exam taken at Rex's Consult, we know that a single CANINE LATERAL THORAX incidence was taken. The Last screen is the image being viewed directly in the browser using a DICOM web viewer.*

It is important to note that the image itself isn't stored in the database. That would have serious implications in performance. As of this moment, the images are being stored in the webroot folder of the web application which isn't the ideal solution either. However, the upload plugin used in the manipulation of the image files is flexible enough to allow for future modifications to the file storage solution allowing even simple integrations with Dropbox and other popular cloud storage solutions or more traditional alternatives. Storing images in the local filesystem of the server limits the ability to scale the application beyond that unique server. An ideal solution would be to store the images on a solution such as Amazon S3 (Amazon Simple Storage Service)[3] and configure it as a CDN[4].

Several plugins with numerous functions were included in the project during development. Search plugins such as the FriendsOfCake Search Plugin made searching and filtering index tables a little bit easier allowing for case-insensitive searches and filters to be applied in the animals, consults and radiography tables. This is especially an area where CakePHP faults and offers very little functionality. File upload is also handled via a 3rd party plugin (josegonzales's cakephp-upload) which made file upload and path storage in the database a slightly more manageable endeavor. And finally the DICOM Web Viewer. An HTML5 and Javascript open source tool for medical image visualization. Using these technologies we can ensure that The imaging files stored in the database can be viewed in virtually any platform that provides a modern browser.

[3] Amazon S3 is a cloud storage service designed to make web-scale computing easier. It has a simple web services interface that can be used to store and retrieve any amount of data. It offers a free tier that would most likely not be enough for the uses of our application but it is competitively priced when compared to other solutions.

[4] A CDN is a globally distributed network of proxy servers deployed in multiple data centers with the goal of serving content to end-users with high availability and high performance serving a large fraction of the internet's content such as live streaming media, graphics, scripts and downloadable objects.

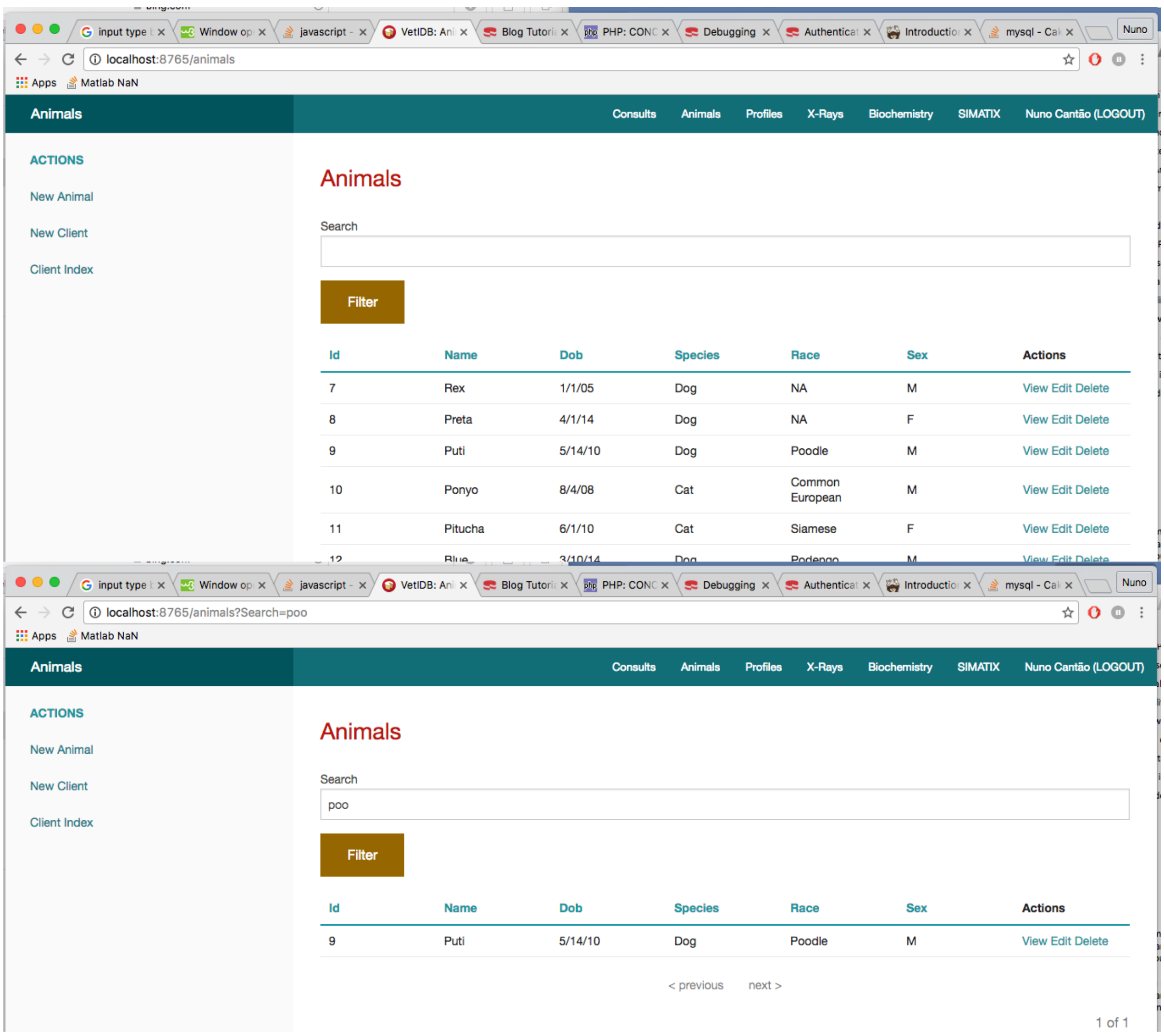



*Figure 21 - Filters/Searches can be made in the Animals/Consults and Radiographies Indexes by the most important fields such as race, species, name, keywords found in report, etc*

# Chapter

# 7 | Conclusions/Future Work

VetIDB, a cloud-based integrated database for animal patients' demographic, biochemical and imaging data, is not only a useful tool for Veterinary practitioners to track their patients' clinical evolution, but also a framework that promotes further advancements in Animal Research, namely the development of new biomarkers.

The exclusive use of simple open source plugins, tools, frameworks and RDBMS was a design choice taken from the very start in order to build the most affordable solution possible for cloud storage in veterinary medicine. All software used in the development of this platform (excluding Sybase's PowerDesigner, whose trial period was used in the beginning of the project for the database design and conception) was handpicked for their capacity of matching (and sometimes even surpassing) other paid-for solutions. The main goal was to build an affordable platform that could be used as an animal's consult report tool whose data could be harnessed in order to further advance both veterinary and human medicine and research.

A lot of work still needs to be conducted in order to turn this into a viable business solution, some of it includes:

1. Improvements on the Design and Responsiveness of the Platform: Although it might be considered as mobile friendly, some components still overlap on each other when viewed on smaller screens and some design work needs to be conducted in order to achieve the status of a true mobile platform;
2. A more permanent and reliable solution for the file storage: Medical Imaging files are quite sizeable by themselves and a platform with the ambitions of VetIDB needs a more reliable and secure solution. Storing the exams in the Server's Filesystem might be a temporary fix but it has severe implications in the scalability of the platform;
3. Better Filtering and Search Tools: Due to time constraints the search tools and options available at the moment, albeit functional and useful still do not meet the requirements elicitation at the beginning of the project. It is possible to

search through animals, imaging exams and consults, although the possibility to apply filters across multiple tables wasn't implemented. This means that, while it is possible to search for animals or exams meeting certain criteria, it isn't possible to cross-filter them automatically according to certain animal characteristics, diagnostics or exam types. That is a job that, at the moment, still has to be done "semi-manually". It is, however important to note the incredible effort that went into linking all information together and guaranteeing easy access to all the related information;

4. Data mining is the computational process of discovering patterns in large data sets involving methods that intersect technologies such as artificial intelligence, machine learning, statistics and database systems. With time, effort and a lot of added functionality, VetIDB could very well become a big contribution in this field for veterinary research.

# References

[1] Smith B. - “The ‘pet effect’–health related aspects of companion animal ownership”. Aust Fam Physician 2012; 41: 439 – 42.

[2] Walsh F. - “Human-animal bonds I: the relational significance of companion animals”. Fam Process 2009; 48: 462 – 80.

[3] Beck AM. - “Review of pets and our mental health: the why, the what, and the how”. Anthrozoos 2005; 18: 441 – 3.

[4] American Veterinary Medical Association - “U.S. Pet Ownership Statistics” [Online] Available at: https://www.avma.org/KB/Resources/Statistics/Pages/Market-research-statistics-US-pet-ownership.aspx [Accessed on August 2016].

[5] Centers for Disease Control and Prevention (CDC) - “Vital signs: prevalence, treatment, and control of hypertension–United States, 1999–2002 and 2005–2008”. MMWR Morb Mortal Wkly Rep 2011; 60: 103 – 8.

[6] Levine GN, Allen K, Braun LT, et al. – “Pet ownership and cardiovascular risk: a scientific statement from the American Heart Association”. Circulation 2013; 127: 2353 – 63.

[7] Hodgson, Kate, et al. "Pets' impact on your patients' health: leveraging benefits and mitigating risk". The Journal of the American Board of Family Medicine 28.4 (2015): 526 – 34.

[8] About the One Health Initiative. [Online] Available at http://www.onehealthinitiative.com/about.php [Accessed on September 2016].

[9] National Research Council (US) Committee on the National Needs for Research in Veterinary Science. Critical Needs for Research in Veterinary Science. Washington (DC): National Academies Press (US); 2005. 1, “The Role of Veterinary Research in Human Society”. [Online] Available at: http://www.ncbi.nlm.nih.gov/books/NBK22905/ [Accessed on August 2016].

[10] World Health Organization’s Health Topics – Zoonoses. [Online] Available at: http://www.who.int/topics/zoonoses/en/ [Accessed on August 2016].

[11] Taylor LH, Latham SM, Woolhouse ME. - “Risk factors for human disease emergence.” Philosophical Transactions of the Royal Society B: Biological Sciences. 2001; 356(1411): 983 - 89.

[12] Müller, Henning, et al. - "A review of content-based image retrieval systems in medical applications—clinical benefits and future directions." International journal of medical informatics 73.1 (2004): 1-23.

[13] Mettler FA, et al. - "Radiologic and Nuclear Medicine Studies in the United States and Worldwide: Frequency, Radiation Dose, and Comparison with Other Radiation Sources 1950–2007". Radiology. 2009 253(2).

[14] Berrington de Gonzalez A, Mahesh M, Kim KP, et al. – "Projected cancer risks from computed tomographic scans performed in the United States in 2007". Arch Intern Med. 2009 Dec 14; 169(22): 2071 – 77.

[15] Brenner DJ, Hall EJ. – "Computed tomography - an increasing source of radiation exposure". N Engl J Med. 2007 Nov 29; 357(22): 2277 – 84.

[14] NDT Resource Center, Iowa State University. - "History of Radiography" [Online] Available at: https://www.nde-ed.org/EducationResources/CommunityCollege/Radiography/Introduction/history.htm [Accessed on August 2016].

[17] Karlsson, Erik B. - "The Nobel Prizes in Physics 1901–2000". Stockholm: The Nobel Foundation. [Online] Available at: http://www.nobelprize.org/nobel_prizes/themes/physics/karlsson/ [Accessed on September 2016].

[18] Frija, J., et al. - "Computed radiography." *Biomedicine & pharmacotherapy* 52.2 (1998): 59 - 63.

[19] Martha Moon Larsen, DVM, MS, DACVR. "Radiographic Interpretation of the normal thorax (Proceedings)" [Online] Available at: http://veterinarycalendar.dvm360.com/radiographic-interpretation-normal-thorax-proceedings [Accessed on September 2016].

[20] The Merriam-Webster Dictionary's entry for "Database" [Online] Available at: http://www.merriam-webster.com/dictionary/database [Accessed on August 2016].

[21] A Relational Database Overview – Oracle [Online] Available at: https://docs.oracle.com/javase/tutorial/jdbc/overview/database.html [Accessed on September 2016].

[22] H. Garcia-Molina, J. D. Ullman, and J. Widom - *“Database Systems : The Complete Book”*, Second Edi. Upper Saddle River, New Jersey: Prentice Hall, 2008.

[23] The Oxford Dictionaries’s entry for "SQL" [Online] Available at: https://en.oxforddictionaries.com/definition/us/SQL [Accessed on September 2016].

[24] IBM’s SQL Guide [Online] Available at: http://www.ibm.com/support/knowledgecenter/SSPK3V_6.3.0/com.ibm.swg.im.soliddb.sql.doc/doc/tables.rows.and.columns.html [Accessed on September 2016].

[25] Gray, Jim (September 1981). - "The Transaction Concept: Virtues and Limitations". *Proceedings of the 7th International Conference on Very Large Databases*. Cupertino, CA: Tandem Computers. [Online] Available at: http://research.microsoft.com/en-us/um/people/gray/papers/theTransactionConcept.pdf. [Accessed on September 2016].

[26] Gray, Jim & Andreas Reuter. *Distributed Transaction Processing: Concepts and Techniques*. Morgan Kaufmann, 1993; ISBN 1-55860-190-2.

[27] Campbell, Robert J. - "Database Design: What HIM Professionals Need to Know." Perspectives in Health Information Management 2004, 1:6 [Accessed on August 2016].

[28] Iitnonline.com The Eclectic History of Medical Imaging [Online] Available at: http://www.itnonline.com/article/eclectic-history-medical-imaging [Accessed on September 2016].

[29] NCBI - National Center for Biotechnology Information [Online] Available at: http://www.ncbi.nlm.nih.gov/ [Accessed on September 2016].

[30] ADNI - Alzheimer’s Disease Neuroimaging Initiative [Online] Available at: http://adni.loni.usc.edu/ [Accessed on September 2016].

[31] Wsj.com Where do you keep all those images? [Online] Available at:
http://www.wsj.com/articles/SB10001424127887323419104578374420820705296
[Accessed on September 2016].

[32] Sherweb.com “Total cost of ownership of servers: how does it add up?” [Online] Available at: http://www.sherweb.com/blog/total-cost-of-ownership-of-servers-iaas-vs-on-premise/ [Accessed on September 2016].

[33] Jan Speelpenning, Patric Daux, Jeff Gallus. – Data Modeling and Relational Database Design – Volume 1, Student Guide. [Online] Available at http://cursuri.cs.pub.ro/~radulescu/bd/er-case/01_00_stud.pdf [Accessed on September

2016].

[34] About QVET [Online] Available at: http://www.qvet.net/Default.aspx [Accessed on September 2016].

[35] PostgreSQL – About page [Online] Available at: https://www.postgresql.org/about/ [Accessed on August 2016].

[35] CakePHP Book – CakepPHP at a Glance [Online] Available at: http://book.cakephp.org/3.0/en/intro.html [Accessed on August 2016].

[36] RegisFray (2008) Wikipedia. [Online] Available at: https://upload.wikimedia.org/wikipedia/commons/a/a0/MVC-Process.svg [Accessed on August 2016].

[37] CakePHP Book – CakePHP Conventions [Online] Available at: http://book.cakephp.org/3.0/en/intro/conventions.html#model-and-database-conventions [Accessed on August 2016].

[38] CakePHP Book – Installation [Online] Available at: http://book.cakephp.org/3.0/en/installation.html [Accessed on August 2016].

[39] Composer - Introduction [Online] Available at: https://getcomposer.org/doc/00-intro.md [Accessed on August 2016].

[40] Amazon Web Services - What is Amazon S3 [Online] Available at: http://docs.aws.amazon.com/AmazonS3/latest/dev/Welcome.html [Accessed on September 2016].

[41] Clifton, Christopher (2010). "Encyclopædia Britannica: Definition of Data Mining". [Online] Available at: https://www.britannica.com/technology/data-mining [Accessed on September 2016].

# Appendixes

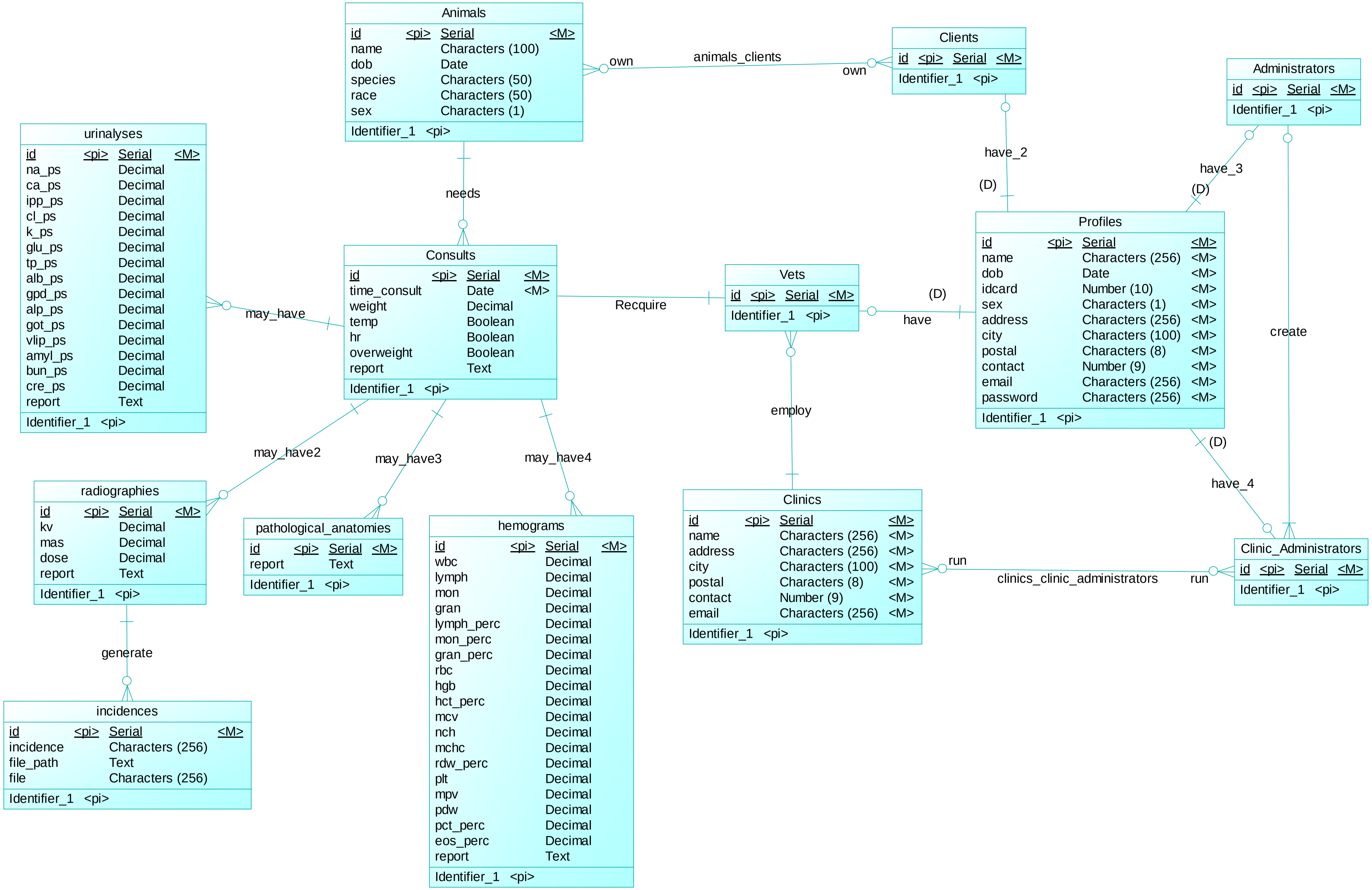



*Figure 22 - Conceptual Model of VetIDB*

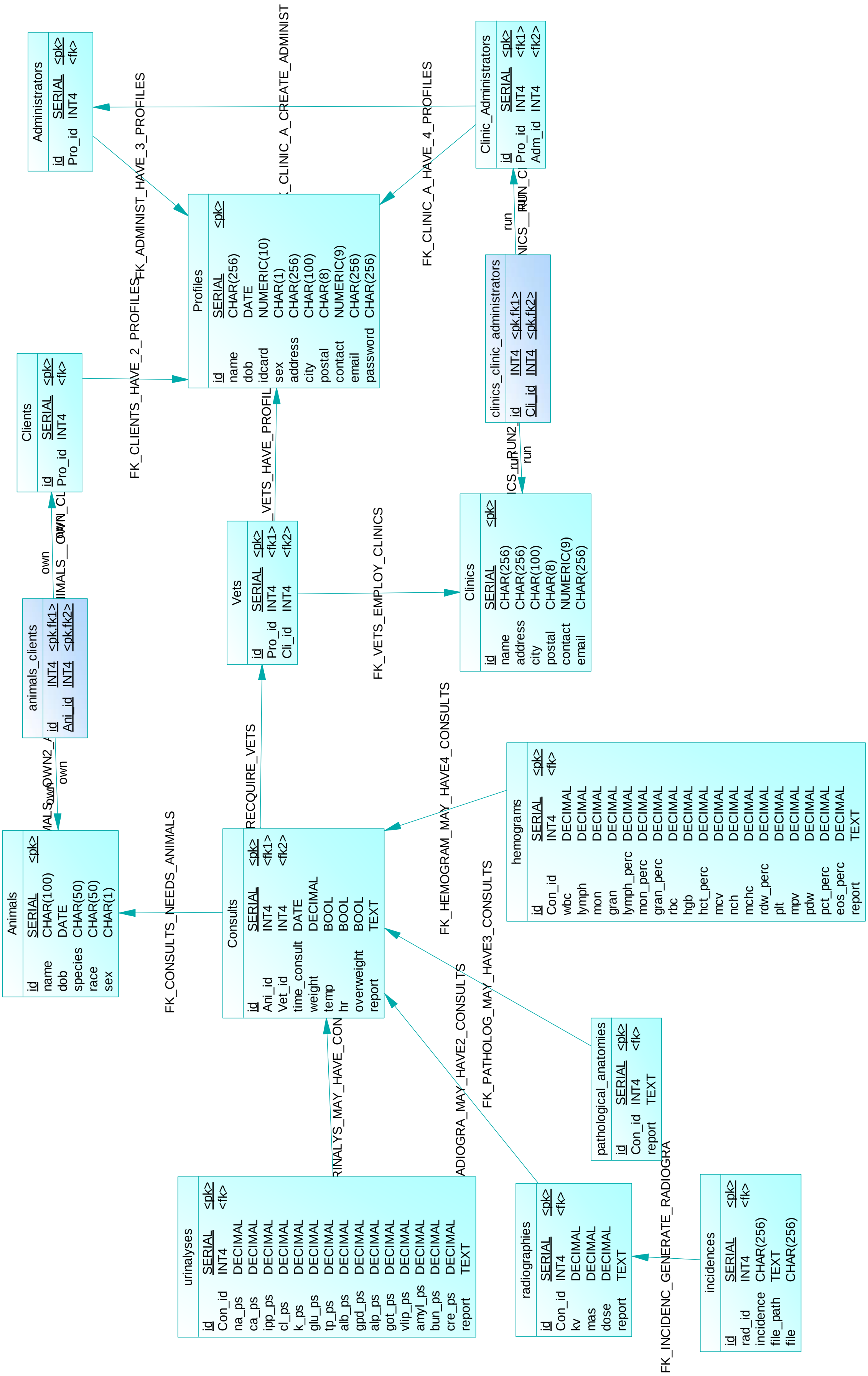



*Figure 23 - Physical Data Model of VetIDB*