Proposed database for the storage and analysis of digital mammograms in Costa Rica

Roberto Mora Gonzlez[1] and Johnny Villalobos Murillo[1]

[1]Imaging Lab, National University, Costa Rica

Introduction

This document contains the description of a database design proposal used to store medical images, in particular, digital mammograms. The following sections contain the relational model, a description of each field as well as implementation requirements.

Hospital System in Costa Rica

"Health care in Costa Rica is mostly public, being primarily managed, financed and provided by two governmental bodies: the Ministry of Health, responsible for governance and regulatory functions, and the Costa Rican Social Security Institute (CCSS), entity in charge of service delivery and financing." (Arocena & Garcia-Prado, 2007, p. 668). The Social Security Institute of Costa Rica currently has 29 public hospitals distributed along all the territory.

There are six hospital types in the public hospital system, these are:

1. National
2. Regional
3. Specialized
4. Peripheral 1
5. Peripheral 2
6. Peripheral 3

"The CCSS offers its services through a network of services organized in three different levels according to its resolutive capacity"(Senz, Acosta, Muiser, & Bermdez, 2011, p. 161). The type of a hospital varies according to its location, the amount of target habitants, the services it offers, among others.

Each of these 29 hospitals has its own director, offered services, attention schedule, and overall norms. More information about each hospital can be found in the Social Security's website[1], as shown in Figure 1.

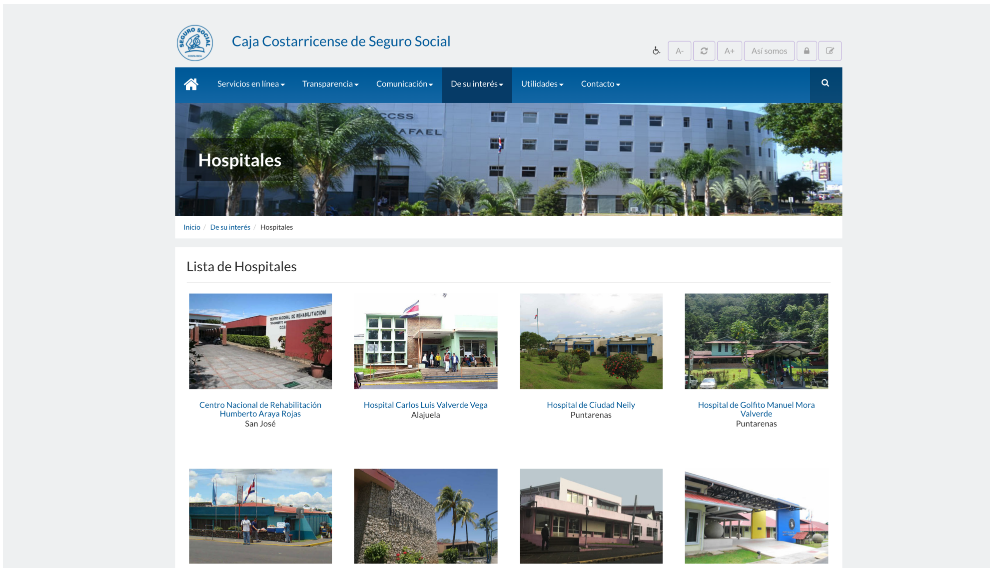

Figure 1: Hospital information found in the CCSS website.
Source: http://www.ccss.sa.cr/hospitales

DICOM overview

The Digital imaging and communications in medicine (DICOM) standard "provides the baseline for the picture archiving and communication systems (PACS)" (Drnasin, Grgi, & Gogi, 2017, p. 538).

DICOM defines both a data format for the imaging information and a network protocol for messaging between devices which implement the standard. DICOM is recognized by the International Organization for Standardization as the ISO 12052 standard.

It also "provides mechanisms to handle data that support cultural requirements, such as different writing systems, character sets, languages, and structures for addresses and person names. It supports the variety of workflows, processes and policies used for biomedical imaging in different geographic regions, medical specialities and local practices" (International Organization for Standardization, 2017). It is commonly said that DICOM is what makes the medical imaging work.

However, this has changed over the years as the Internet based applications have demanded access to PACS without using the DICOM protocol. This has lead to the development of Web Access to DICOM Objects (WADO). "WADO allows access to DICOM objects on web-enabled DICOM server through the Hypertext Transfer Protocol/Secure (HTTP/S) protocol using DICOM unique identifiers (UIDs)" (Drnasin et al., 2017, p. 538).

[1] http://www.ccss.sa.cr/hospitales

This is why this paper proposes a data model inspired in the DICOM standard, but implemented with a traditional object relational database management system. This allows for the creation of systems that access this database over a server on the web, which adapts more closely to the rapidly moving internet.

Figure 2, shows the base of the DICOM model entities which were used as center for the creation of the proposed relational model. Note how the DICOM standard defines a one-to-n relationship between its core nodes, from patient up until the instances themselves.

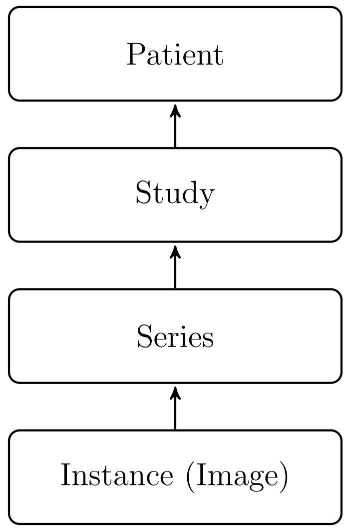

Figure 2: Basis of the DICOM model.

Proposed database model

The database model derived from the DICOM standard has three logic areas. The first one contemplates the storage of the medical imaging by implementing a model based on the DICOM standard data model. The second area adds analysis to a medical imaging database by incorporating information about regions found in the segmented medical images. This regions can be later used to perform morphological analysis using SQL queries.The third area adds information about the hospitals related to the imaging data. This area was inspired in the Costa Rican hospital system.

Figure 3, shows the relational model for the medical imaging database. Note how the tables are grouped into the aforementioned logic groups.

This particular database designs was created for PostgreSQL using some of its features as an object-relational database management system (ORDBMS). This is why there are some fields which are arrays and also why there are some composites as data types for some fields in the Instance and Region tables. In PostgreSQL, "a composite type is simply a list of types with associated field names. A value of a composite type is a row or record of field values. The user can access the component fields from SQL queries"(The PostgreSQL Global Development Group, 2018, p. 1024).

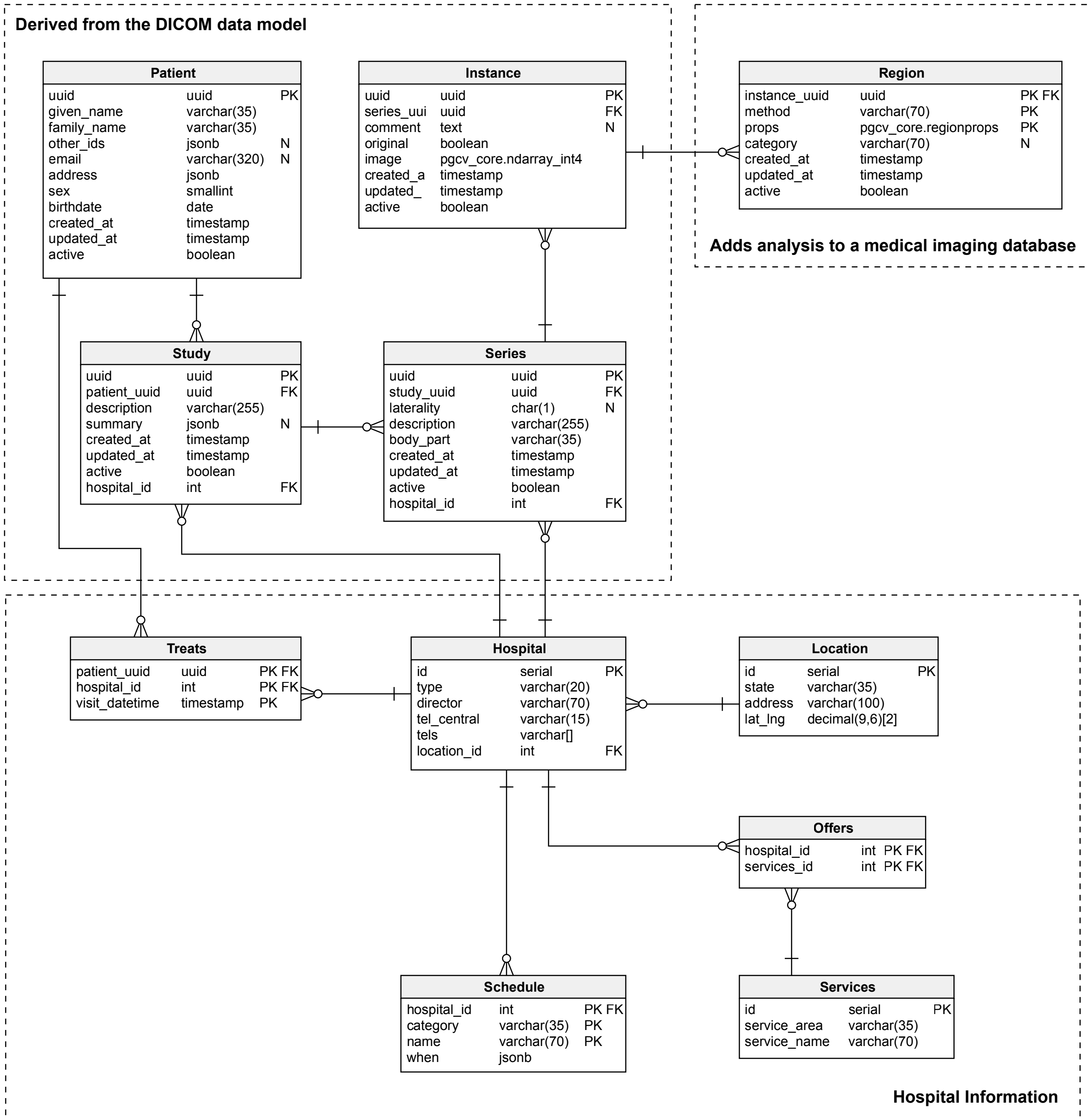

Figure 3: Relational model proposed for the medical imaging application.

The following are the descriptions for each of the tables in this data model. Theese tables include the column names, the column types (according to the types supported by PostgreSQL or the types defined within *pgcv*), the column properties and the description for each column in the tables.

Table Patient			
Column name	**Type**	**Properties**	**Description**
uuid	uuid	PK	Universally unique identifier of the patient in the database.
given_name	varchar(35)		The given name of the person.
family_name	varchar(35)		The family name of the person, usually the last name.
other_ids	jsonb	null	JSON array of alternate identifiers for the patient.
email	varchar(320)	null	Email address of the patient
address	jsonb		JSON object containing information about the region of the patient, for instance, country, state/province, city
sex	smallint		Sex of the patient, according to the ISO/IEC 5218: Codes for the representation of human sexes (International Organization for Standardization, 2004).
birthdate	date		Birthdate of the patient
created_at	timestamp		Date and time when the patient was created
updated_at	timestamp		Date and time when the patient was last updated
active	boolean		Defines whether the patient is active

Table 1: This table stores basic patient information inspired by the DICOM standard.

Table Study			
Column name	**Type**	**Properties**	**Description**
uuid	uuid	PK	Universally unique identifier of the study in the database.
patient_uuid	uuid		Reference to the uuid of the patient.
description	varchar(255)		Description or classification of the Study performed.
summary	jsonb	null	Stores information about the overall study and reporting information, for instance, the BI-RADS category
created_at	timestamp		Date and time when the study started
updated_at	timestamp		Date and time when the study was last updated
active	boolean		Defines whether the study is active
hospital_id	int		Hospital where the study started

Table 2: This table stores basic study information inspired by the DICOM standard.

Table Series			
Column name	**Type**	**Properties**	**Description**
uuid	uuid	PK	Universally unique identifier of the series in the database.
study_uuid	uuid		Reference to the uuid of the study.
laterality	char(1)	null	Laterality of (paired) body part examined. Required if the body part examined is a paired structure. R = right, L = left
description	varchar(255)		User provided description of the Series
body_part	varchar(35)		Text description of the part of the body examined
created_at	timestamp		Date and time when the series started
updated_at	timestamp		Date and time when the series was last updated
active	boolean		Defines whether the study is active
hospital_id	int		Hospital where the series were captured

Table 3: This table stores basic series information inspired by the DICOM standard.

Table Instance			
Column name	**Type**	**Properties**	**Description**
uuid	uuid	PK	Universally unique identifier of the instance in the database.
series_uuid	uuid		Reference to the uuid of the series.
comment	text	null	
original	boolean		Describes whether an image pixel values were based on source data or have been derived in some manner from the pixel value of one or more other images
image	pgcv_core.ndarray_int4		The datatype containing the image data. Refer to the pgcv PostgreSQL extension.
created_at	timestamp		Datetime the image pixel data creation started.
updated_at	timestamp		Datetime the image pixel data was last updated
active	boolean		Defines whether an instance is active

Table 4: This table stores basic instance/image information inspired by the DICOM standard.

Table Region			
Column name	**Type**	**Properties**	**Description**
instance_uuid	uuid	PK	Reference to the uuid of the instance.
method	varchar(70)	PK	Method used to extract the region properties. This allows for different segmentation functions
props	pgcv_core.regionprops	PK	The datatype containing properties of the region. Refer to the pgcv PostgreSQL extension.
category	varchar(70)	null	Class of region according to a classifier
created_at	timestamp		Datetime the region information was extracted from the image
updated_at	timestamp		Datetime the region was last updated
active	boolean		Defines whether the region is active

Table 5: This table stores information about the regions found in an image using a segmentation method.

Table Hospital			
Column name	**Type**	**Properties**	**Description**
id	serial	PK	The hospital identifier and primary key
type	varchar(20)		The type of hospital. In Costa Rica, there are 6 types of hospitals
director	varchar(70)		The name of the hospital's director
tel_central	varchar(15)		Telephonic central of the hospital
tels	varchar[]		Telephone numbers of the hospital
location_id	int		The identifier of the hospital location tuple

Table 6: Table that stores information about the hospitals.

Table Location			
Column name	**Type**	**Properties**	**Description**
id	serial	PK	The identifier of each location tuple
state	varchar(35)		The state/province where the hospital is located
address	varchar(100)		The given street address of the hospital
lat_lng	decimal(9,6)[2]		A latitude and longitude tuple indicating the hospital's coordinates

Table 7: Location information of the hospitals.

Table Services			
Column name	**Type**	**Properties**	**Description**
id	serial	PK	The identifier of the service
service_area	varchar(35)		The service area that includes particular services
service_name	varchar(70)		The given service name

Table 8: Services of the hospitals.

Table Offers			
Column name	**Type**	**Properties**	**Description**
hospital_id	int	PK	Identifier of the hospital that offers the service
services_id	int	PK	Identifier of said service

Table 9: Relationship that stores which services are offered by which hospitals.

Table Schedule			
Column name	**Type**	**Properties**	**Description**
hospital_id	int	PK	Hospital referencing the schedule
category	varchar(35)	PK	Category of the schedule, e.g.: visits
name	varchar(70)	PK	Name of the schedule
when	jsonb		JSON array containing schedule objects. Each schedule objects contains a list of days and a time interval

Table 10: Attention schedules of each hospital.

Table Treats			
Column name	**Type**	**Properties**	**Description**
patient_uuid	uuid	PK	The patient identifier
hospital_id	int	PK	The hospital identifier
visit_datetime	timestamp	PK	A timestamp of when the patient was treated

Table 11: Table that registers every-time a patient visits an hospital, meaning, every time a patient is treated in a given hospital.

Analysis tool

As mentioned in the previous section, the proposed database model incorporates some of the facilities of object-relational database management systems (ORDBMS), more specifically, the addition of specialized composite types. This is done this way, because "the object-relational data model extends the relational data model by providing a richer type system including complex data types and object orientation" (Silberschatz, Korth, & Sudarshan, 2011).

The mentioned composite data types are taken from the *pgcv PostgreSQL extension*, which is an extension that implements data types and functions for computer vision from the database server. This extension harvests the maturity of the scientific Python ecosystem, incorporating SQL functions whose body is written in Python using PL/Python, which is a procedural language for PostgreSQL.

Figure 4, shows the schemas included into the *pgcv* extension. These modules include functionally similar functions for image manipulation. It's worth noting that while *pgcv* does not perform the mammogram analysis by itself, it provides the necesary tools to accomplish it, by implementing image manipulation functions.

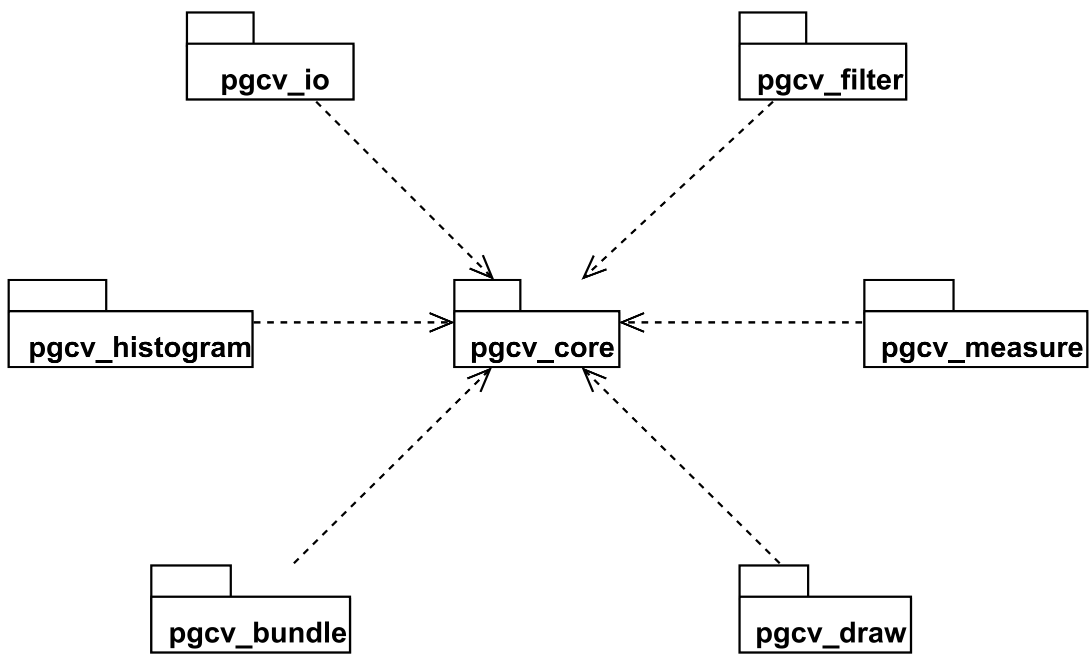

Figure 4: Schemas implemented in *pgcv*. This schemas group functions into functionally similar groups

More information about this extension can be found in the Imaging Lab website[2] or in the GitHub code repository[3].

[2] http://imaginglab.una.ac.cr/index.php/rf01/p051116/p2
[3] https://github.com/romogo17/pgcv

Conclusion

After analyzing the hospital system in Costa Rica, it is possible to observe how, using the proposed relational model, incorporating medical imaging analysis to the current infrastructure is a straightforward process.

Furthermore, it is possible to use a database extension that allows the manipulation of digital mammograms. This extension can be used to create a variety of applications, such as the database model proposed by this paper or a graphical user interface (GUI) that query and execute image manipulation functions.

In fact, as part of the contributions made by the National University's Imaging Lab, there is also a support tool that works as a proof of concept to this idea. This tool, allows the user to navigate the medical imaging data model derived from the DICOM standard as well as the regions found in a given image. Even though, this support tool was designed primarily for mammogram analysis, its features could be further improved in order to add other areas as well. More information about this tool can be found in the Imaging Lab website[4].

This way, it is possible to harvest the object relational database management systems (ORDBMS) potential along with the mature state of the scientific Python ecosystem in order to create a applications that would potentially serve as contributions to the national health system.

References

Arocena, P., & Garcia-Prado, A. (2007, July). Accounting for Quality in the Measurement of Hospital Performance: Evidence from Costa Rica. *Health Economics*, *16*(7), 667–685.

Drnasin, I., Grgi, M., & Gogi, G. (2017, October). JavaScript Access to DICOM Network and Objects in Web Browser. *Journal of Digital Imaging*, *30*(5), 537–546. doi: 10.1007/s10278-017-9956-7

International Organization for Standardization. (2004). *Information technology – Codes for the representation of human sexes.*

International Organization for Standardization. (2017). *Health informatics – Digital imaging and communication in medicine (DICOM) including workflow and data management.*

Silberschatz, A., Korth, H. F., & Sudarshan, S. (2011). *Database system concepts* (6th ed ed.). New York: McGraw-Hill. (OCLC: ocn436031093)

Senz, M. d. R., Acosta, M., Muiser, J., & Bermdez, J. L. (2011, November). Sistema de salud de Costa rica. *The health system of Costa Rica.*, *53*, S156–S167.

The PostgreSQL Global Development Group. (2018). *PostgreSQL 10.4 Documentation.* PostgreSQL Global Development Group.

[4] http://imaginglab.una.ac.cr/index.php/rf01/p051116/p3