

# Data Pipeline Assignment

Nelson López

June 2021

## Project Report

### 1. Introduction

The goal of this project is to detail all the process that went necessary to build XML data base model for BigVir2018, as well as the proposed visualization scenarios taking in consideration data restriction policies for each visualization.

#### 1.1 Assignment

The BigVir2084 application is intended to enable public health monitoring and contact tracing by Government instances on BigVir-related data to alert citizens on the BigVir epidemic and infection risks and make Government decisions on the isolation of infected or suspected citizens, the population lockdown, the opening conditions, or closure of establishments receiving the public, and the public or private gathering conditions or prohibitions as well.

The BigVir2084 database stores general personal information on citizens, like age and weight, in addition to medical history: consultations, hospitalizations, laboratory test results, etc., and, most importantly infection, immunization and vaccination statutes, and finally social relations and social history: presence in public establishments, participation in events, etc.

Citizens can access their statutes, and those of events, get alerts and general information. Licensed physicians can update the citizens' medical information and status, the global health observatory can access all the information on citizens, establishments, and events to perform data analytics and compute health indicators, and the Government agency can update the status of any BigVir2084 agent according to health indicators and health policies.

## 2. Design

About the fields of the required data base modelling, I chose the following structure:

**For general information:**

Last name, First Name, Date of Birth, Weight, Height, Address, Zip Code, Phone Number, Nationality, Email Address.

**Explanation:** Storing this type of data, allows us to visualize citizen's location and do cross data between medical history and social relations.

**For Medical information and history:**

**Consultations:** Doctor, Consultation's date, Doctor specialty, Consultation's code (severity of consultation), Consultation's city and Consultation's country.

**Hospitalization:** Hospital, Hospital city, Hospital Country, Hospitalization's date, Specialty consulted, Severity code.

**Laboratory:** Test type, Test date, Test date, Test city, Test country, Test Result

**Vaccination Status:** Vaccine Type, Vaccination date, Vaccination city, Vaccination Country, Vaccinated (Boolean value)

**Explanation:** Storing this data is useful when we analyze the impact of a vaccine or a health problem like covid19, we will have the possibility to identify if a vaccine is being effective or no, if we have a common problem in a specific location, many possibilities of cross data, we have an c\_ID attribute that allows us to cross all this data by citizen.

**For Social Relation's information and history:**

**Event:** Event Type, Event name, Event city, Event country, Event date

**Explanation:** Same as Medical information, this data allows to cross medical, location and social relations data, all each citizen. The use case may be when identified a common health problem, we can cross the participation of a specific infected citizen with another non infected citizen and try to prevent the spreading of the infection calling them.

**DB - Doctors:** Storing doctor's info by ID

**Fields:** Doctor last name, Doctor First name, Doctor specialty

**Explanation:** Doctor's data base with ID that we use inside citizens information, allows us to identify common patients, common doctors, and many possibilities of visualization.

**DB - Vaccines:** Storing vaccine's info by ID

**Fields:** Vaccine name, Vaccine info, Vaccine dose (number of the corresponding dose)

**Explanation:** same as Doctor's data base, we are using Vaccine Attribute to crossing data versus citizens and do some statistical measurements of effectiveness of each one.

### 3. Visualization scenarios

#### Introduction:

XSLT scenarios were built in two ways, **scenarios 1 and 3** were built without recursive programming, **scenarios 2 and 4** were built with recursive programming, this way we can use all knowledge acquired in class.

#### Scenario 1:

Scenario created to give information of a specified citizen, allowing him/her to view all his/her information storage. The specified citizen will not be allowed to visualize another information, Citizen ID represents something like a credential, social security number, etc. that it is a unique identifier.

#### Scenario 2:

Scenario created to do some computations and exploit the database, using recursive templates!

#### Scenario 3:

Scenario created for Global Health Observatory, this organization can evaluate all data to do analysis and prevent, predict, or identify a massive health problem

#### Scenario 4:

Scenario created to proportionate a visualization general for Doctor's where they can visualize their patients and some general information like name, last name and phone number, a transforming data to a new XML structure to exploit.

### 3. Tools used

I used the following tools to complete the modelling

- Visual Studio 2019
- XSLT 1.0
- XML 1.0

