Canreg database processor

# Introduction

CanReg Database Processor (CRDP) is a tool built in Java and R to help CanReg4 users to migrate their database to CanReg5. The tool will generate 3 files (patient.csv, tumour.csv, source.csv) ready to be imported into CanReg5.

## What do we need to install the crdp

First you have to download the CanReg Database Processor from this link:

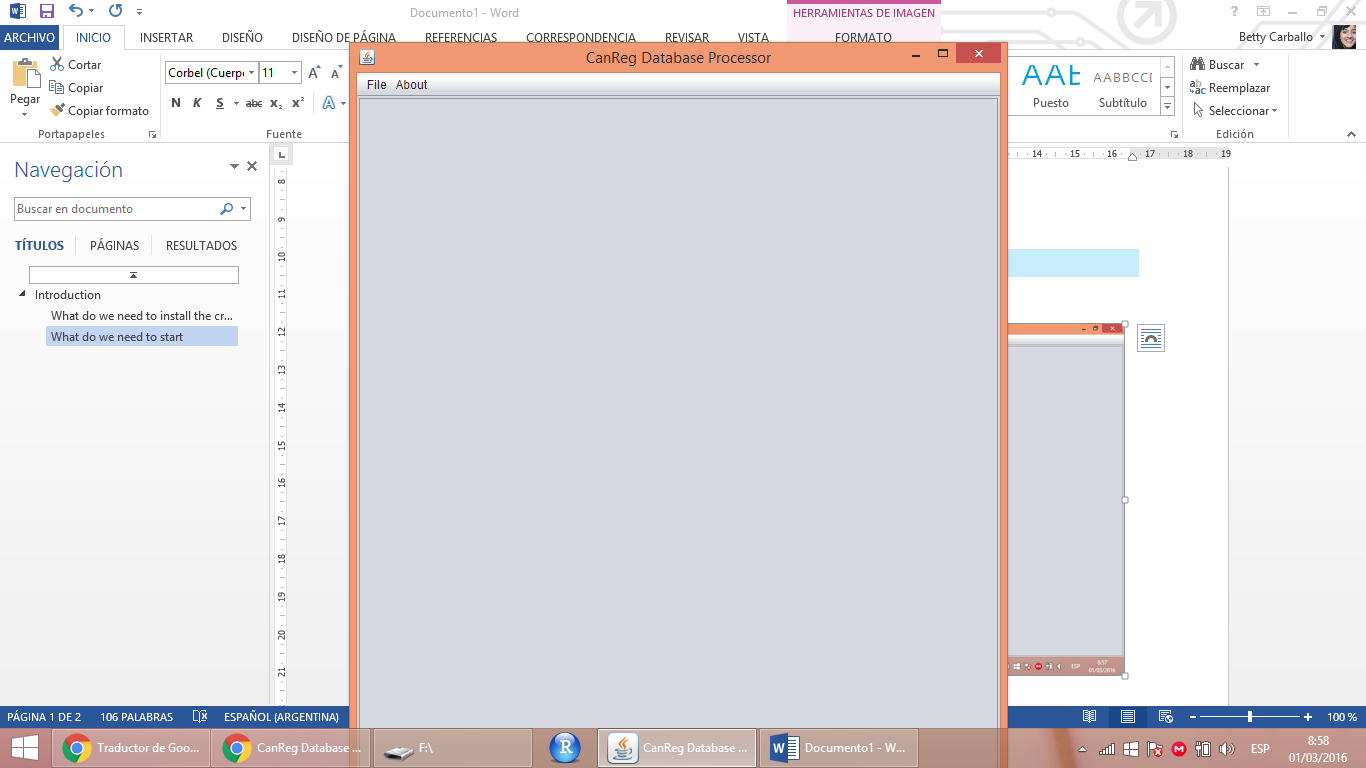
<https://drive.google.com/file/d/0B97PBG5R1R9VckdINm9ESW9zOWc/view?usp=sharing>

It’s a little bit heavy but it has Java, R and other tools necesary for the CRDP to run.

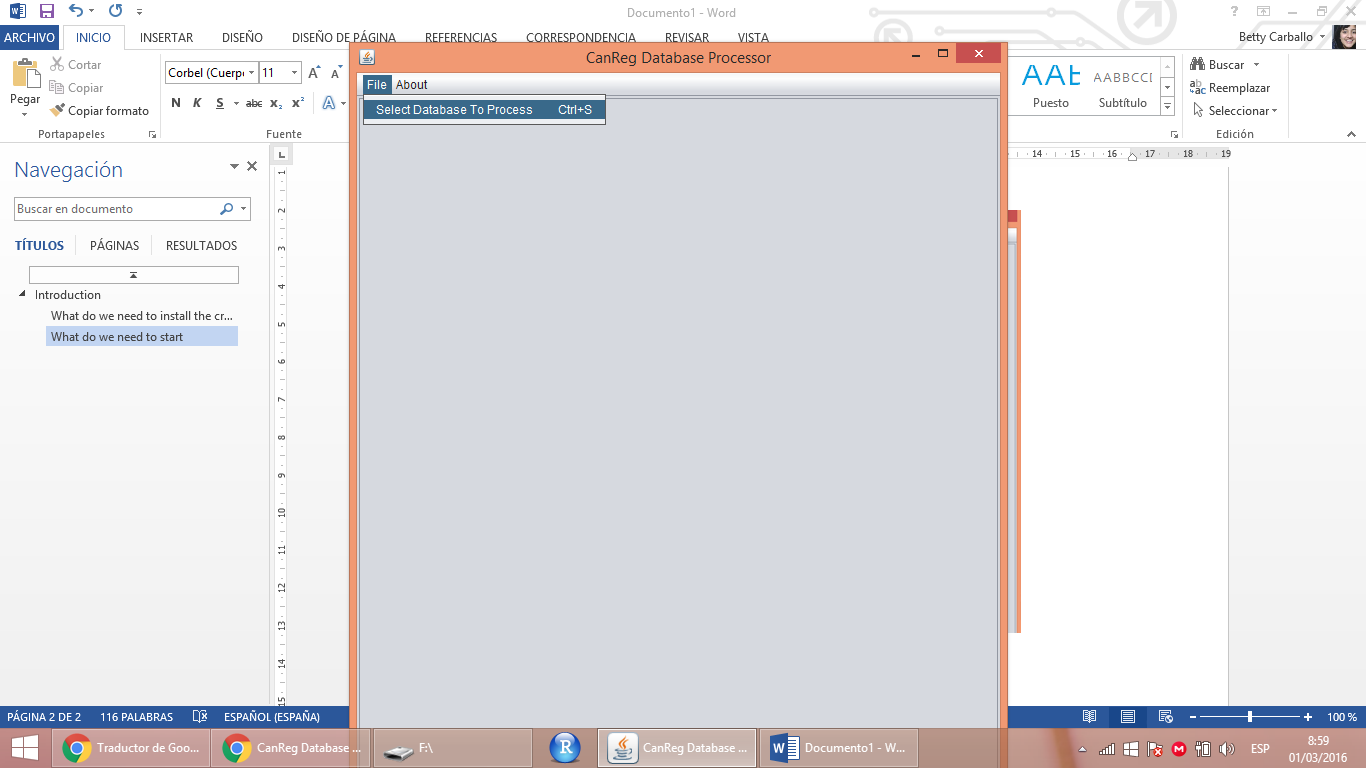
The installation is a straightforward process, just click next and accept everything.

## Steps

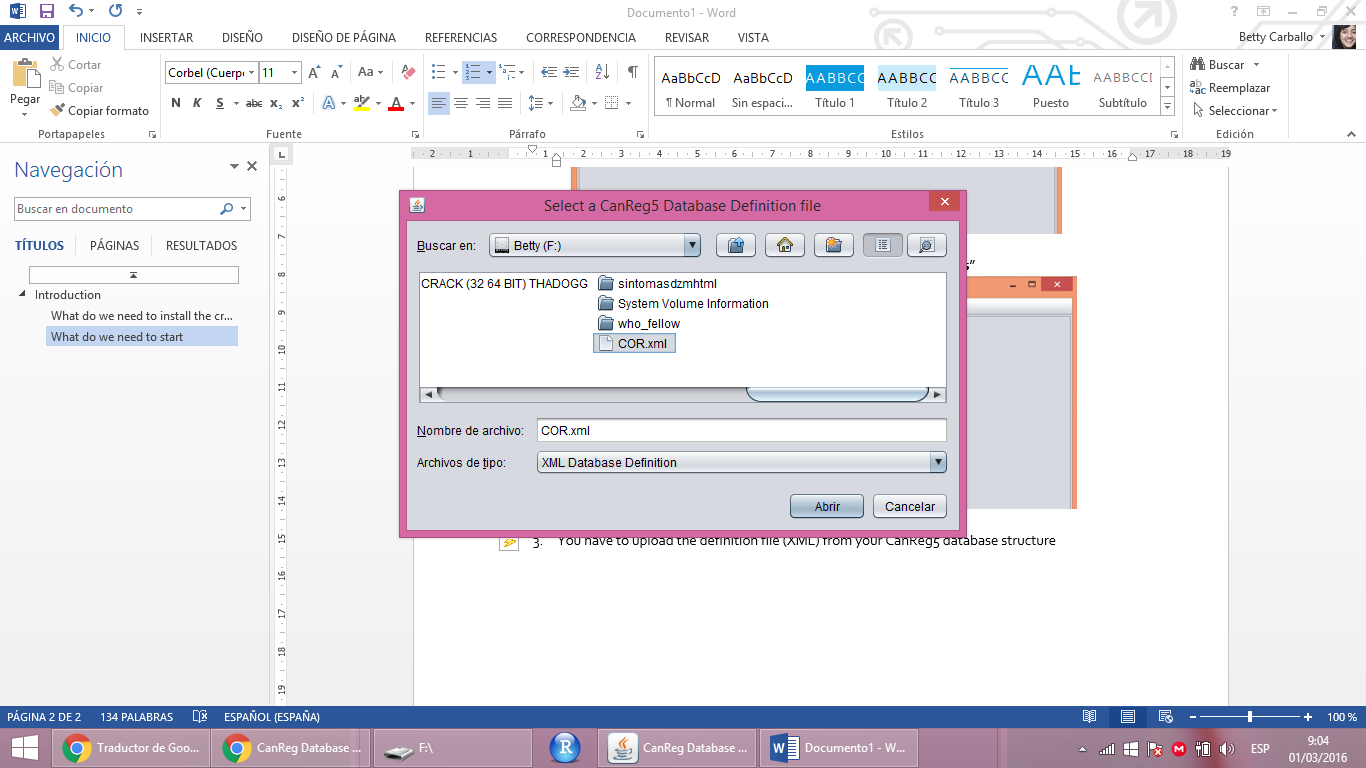
1. Run the CRDP



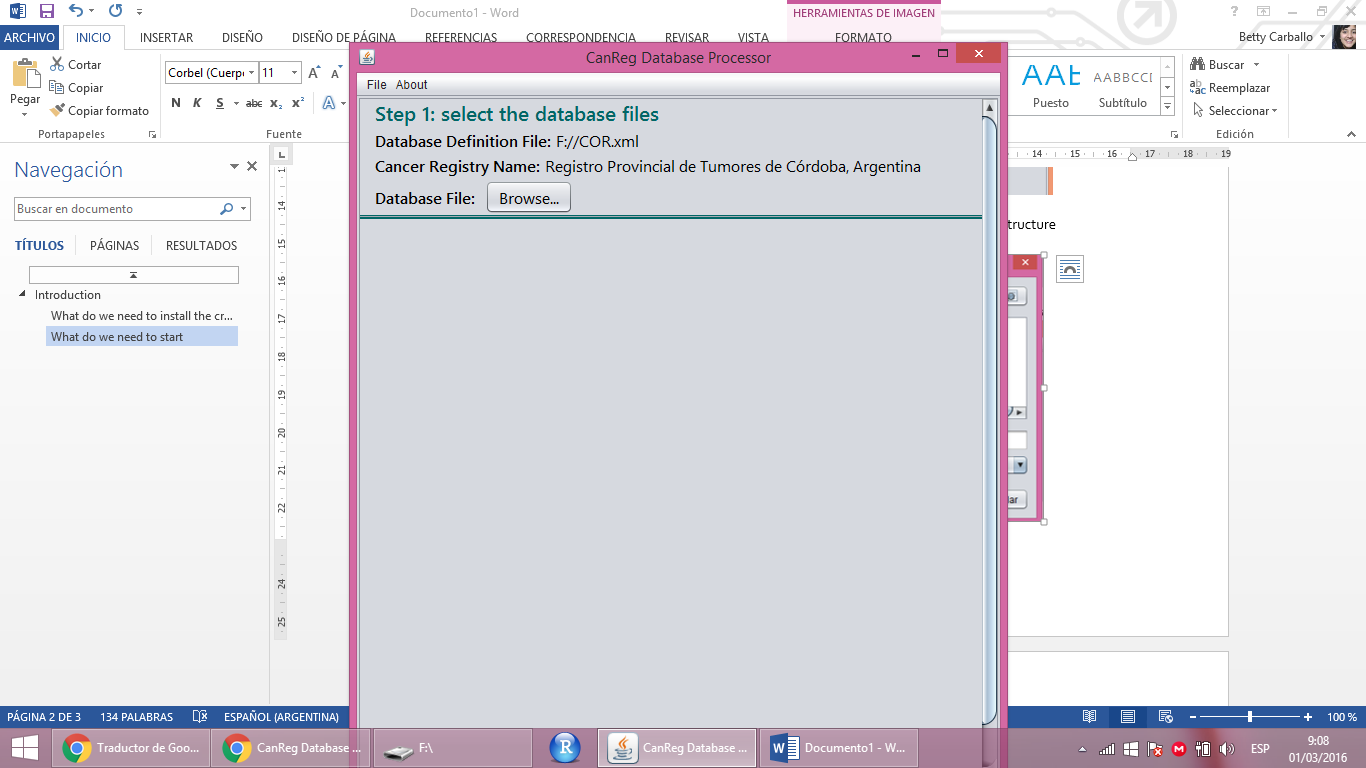
1. Go to file and you have just one option: “**Select database to process**”



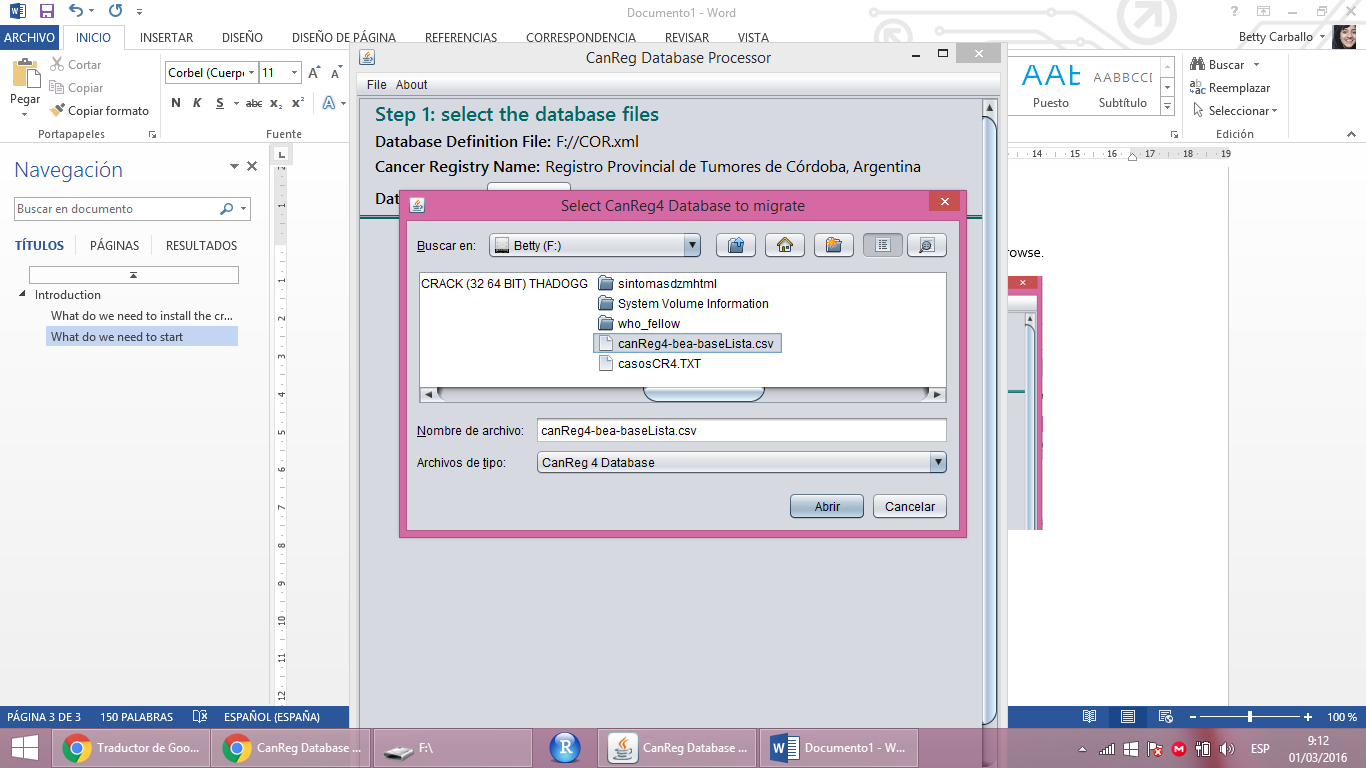
1. You have to upload the definition file (XML) from your CanReg5 database structure



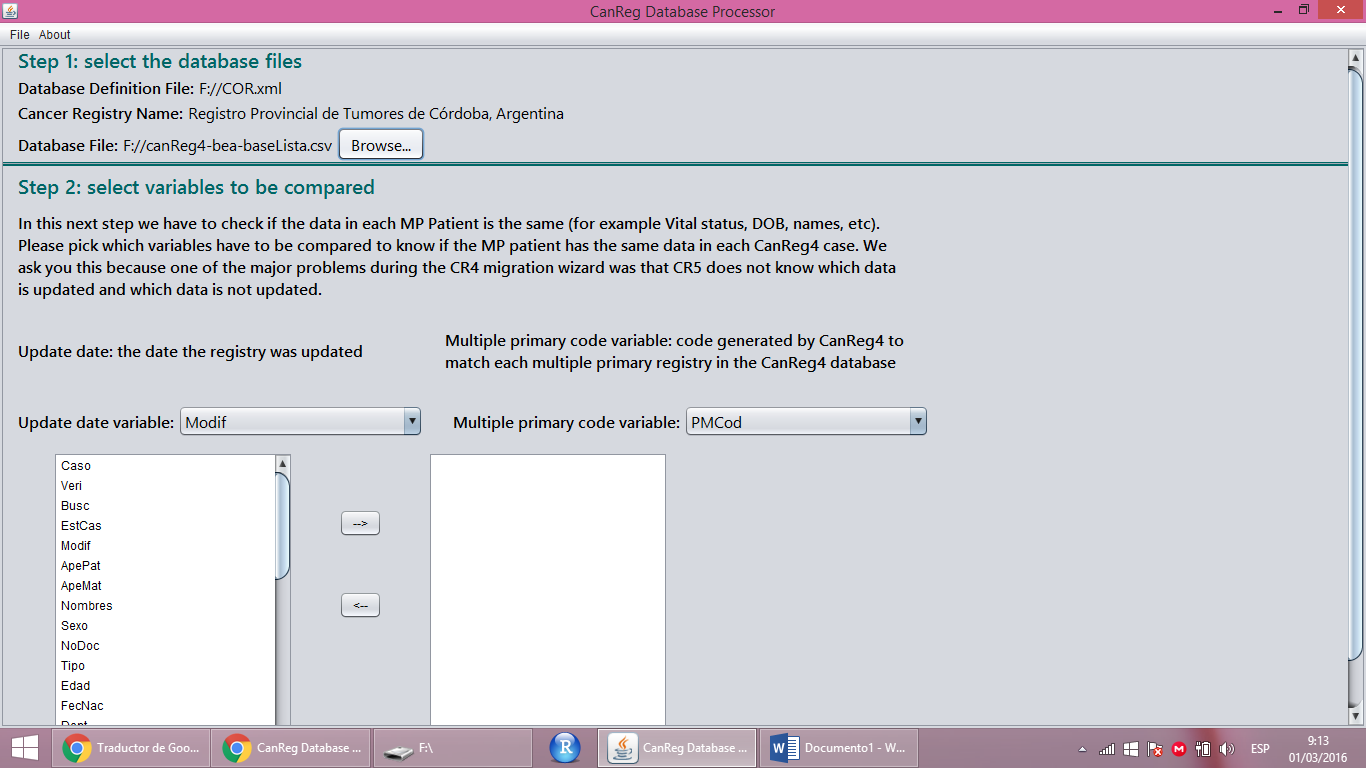
1. Now you have to select the database that you want to migrate. Click on Browse.

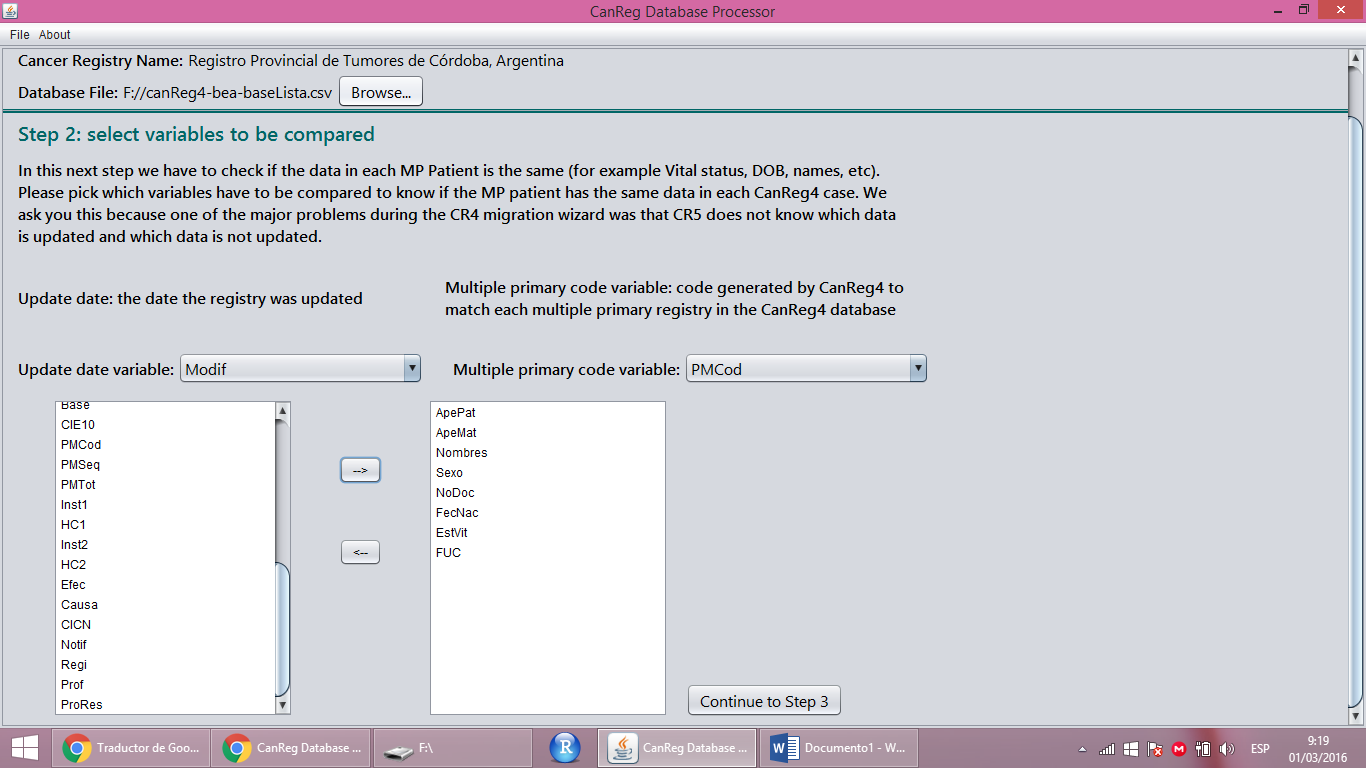


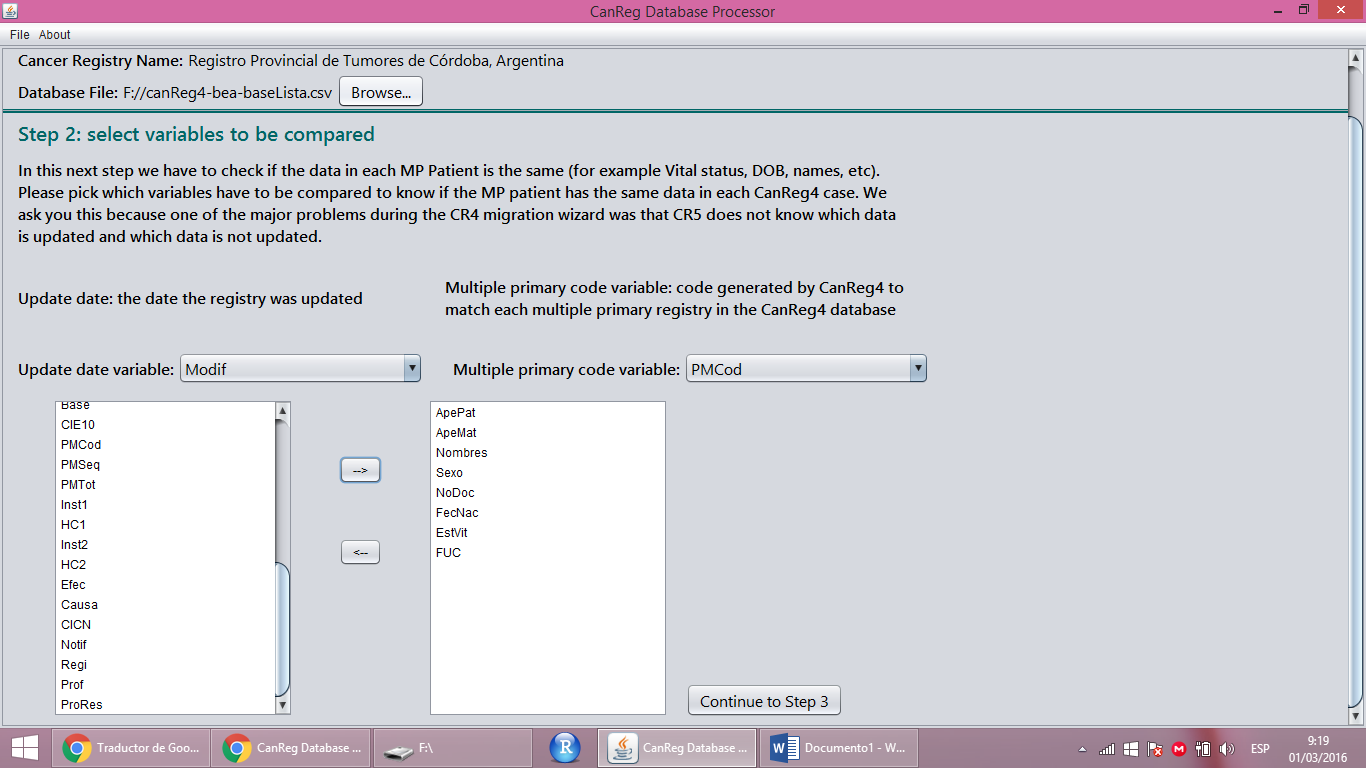
**Note: the database file has to be comma separated values.**



1. The next step is to select the multiple primary patient variables that have to be compared. In CanReg5 the wizard does not know which data for each patient is correct, so if the same patient has different data (e.g. John Smith DOB: 01/07/1968 and John J. Smith DOB 5/6/1968) the wizard will keep both patients even though is the same one.

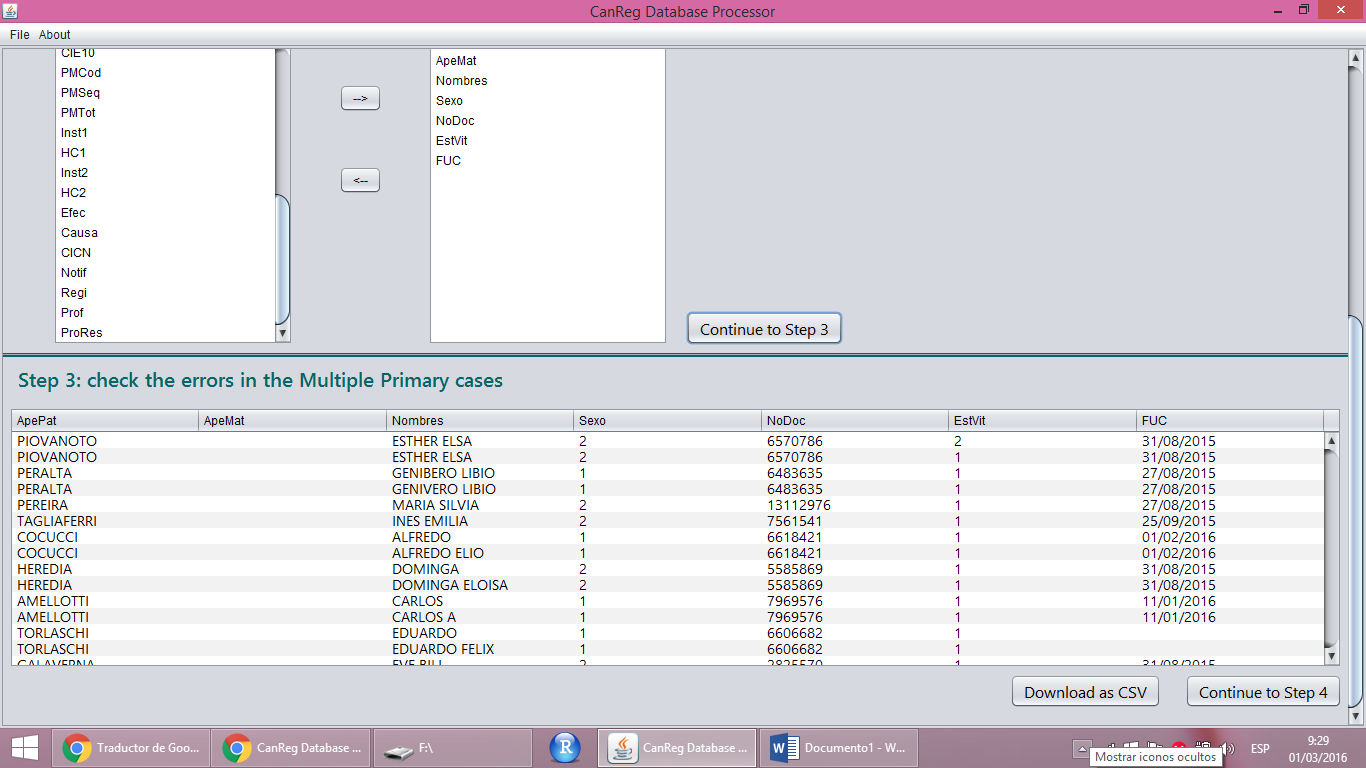


1. After we select the “Update Date” variable, the “Multiple Primary Code” variable and all the variables to be compared we have to click 

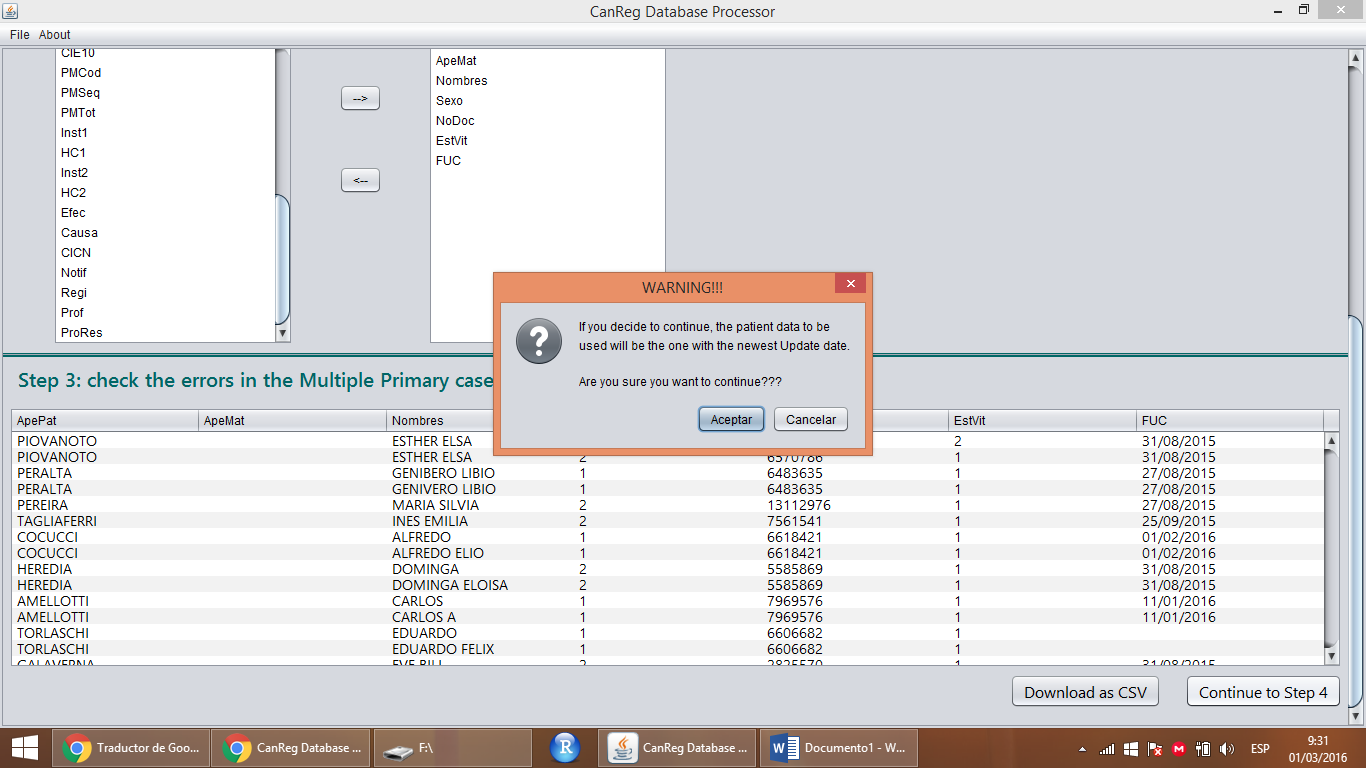


1. Note: after clicking in “Continue to step 3” we have to scroll down. This is a bug that the scroll bar does not go down automatically.

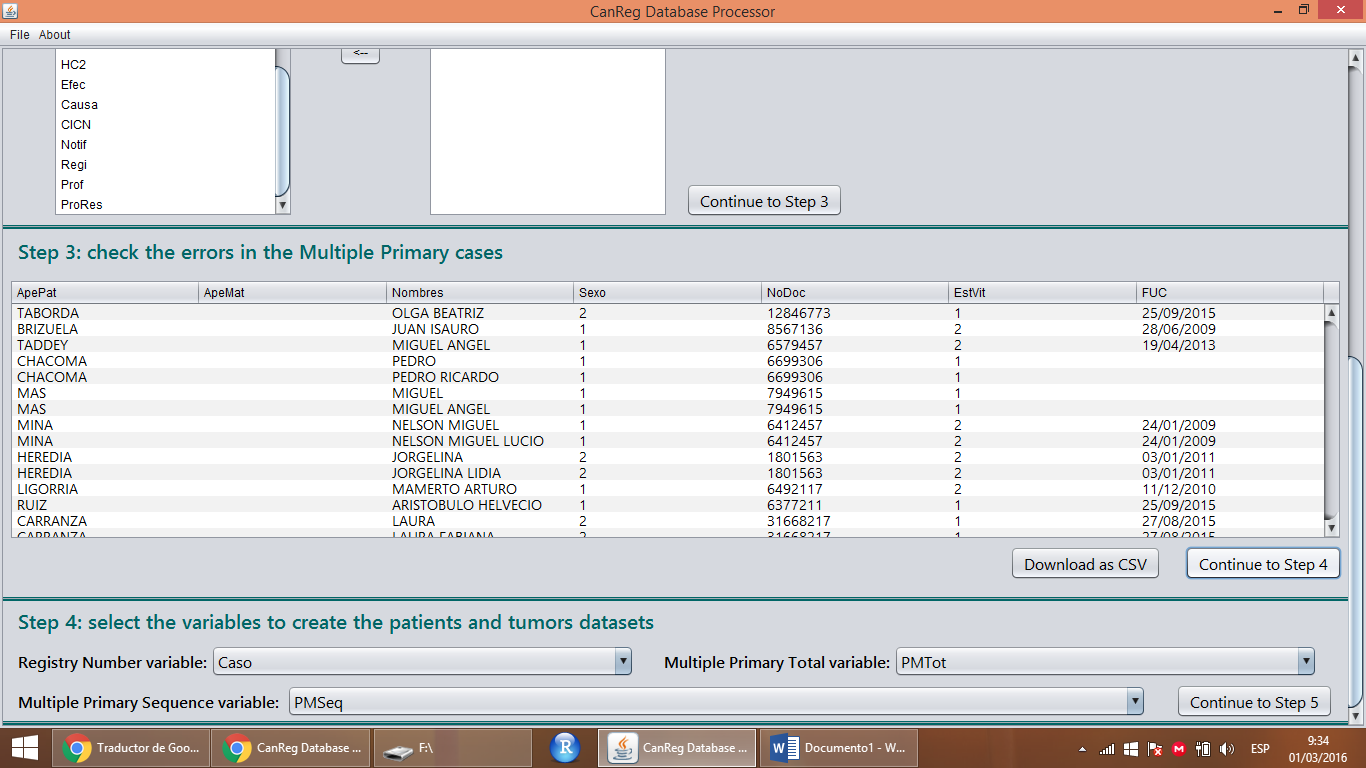
A table with the Multiple Primary Patients that do not have the same information will appear.



The table with the errors can be downloaded, so you can fix the errors and start the process again. Or you can Continue to Step 4, but the newest patient is going to be kept (a warning message is going to be prompted).



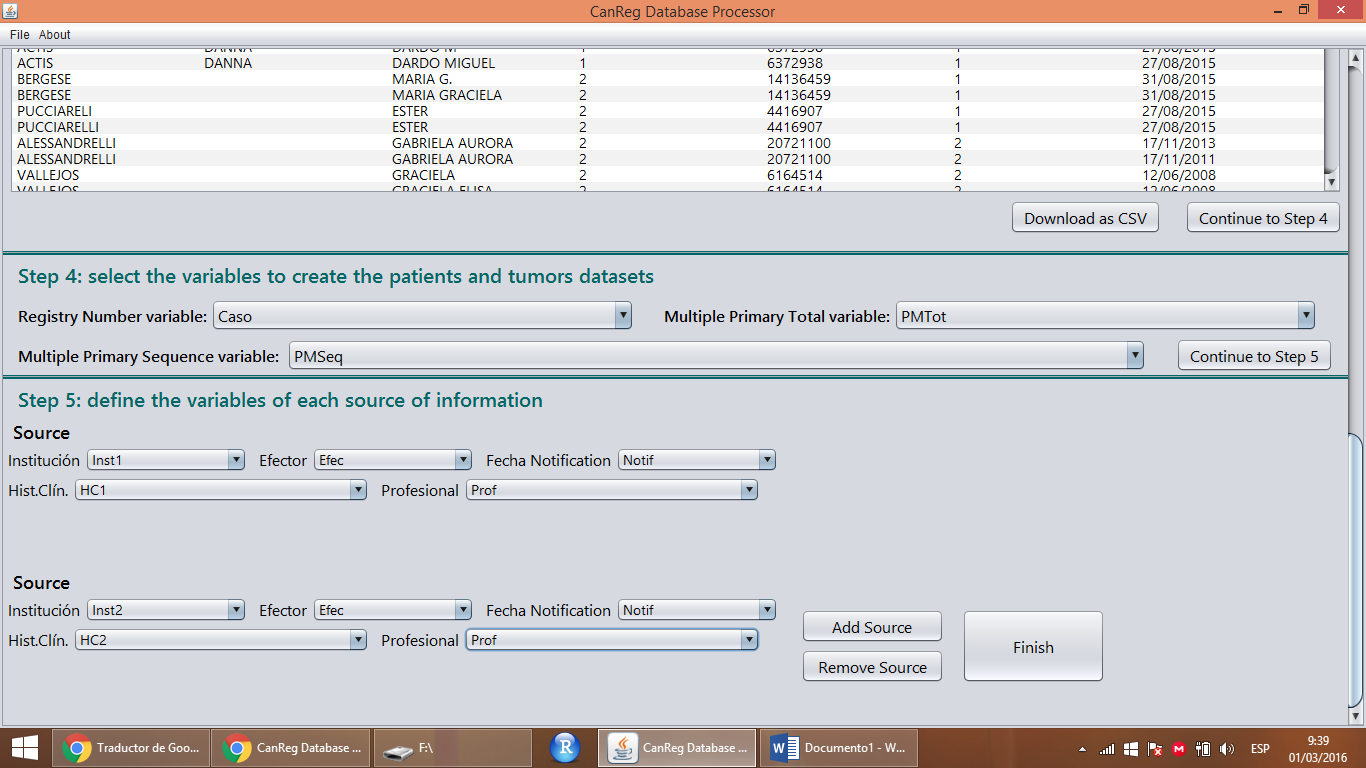
1. The user has to select a few more variables to create the dataset for patients and tumours: Registry Number, Multiple PrimaryT and Multiple Primary Sequence. Then click on “Continue to Step 5” (the last one).



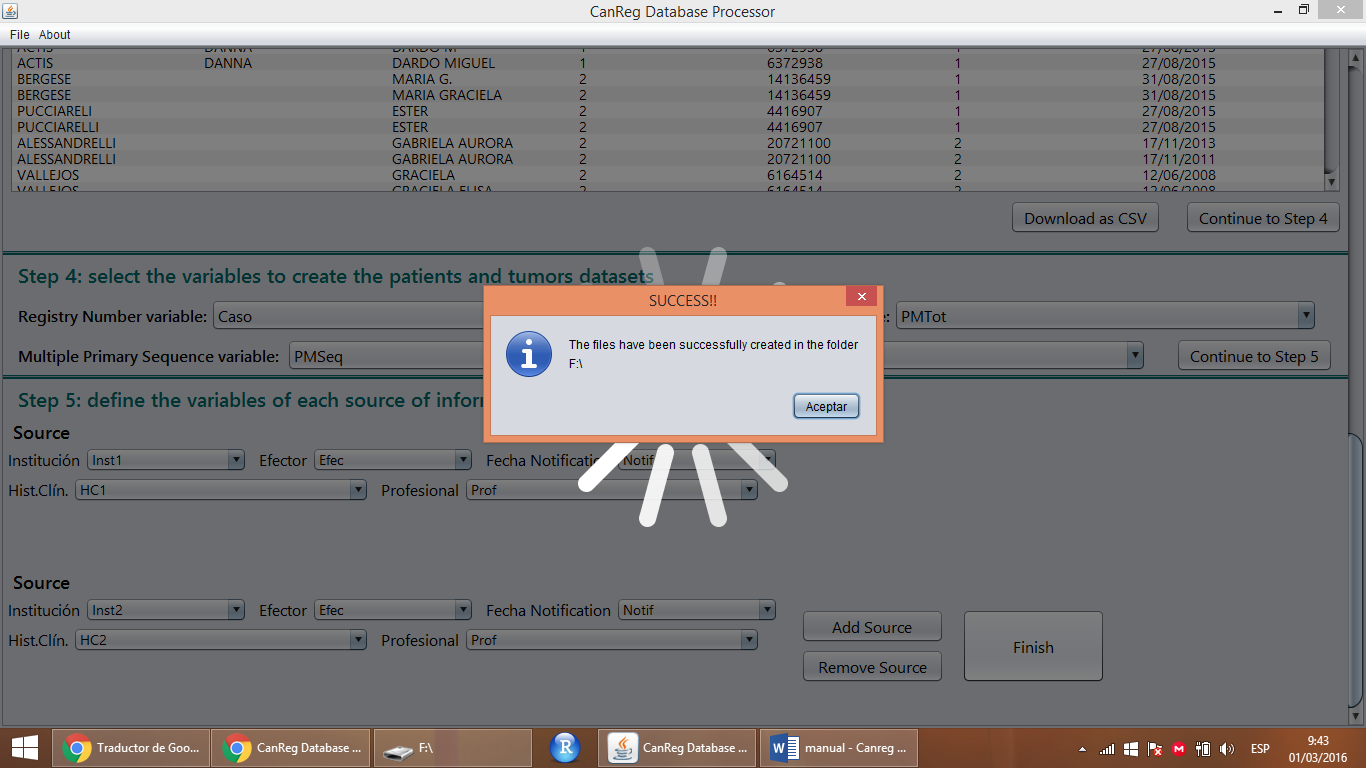
1. The last step is selecting the columns for each source. In CanReg5 for each source there has to be one row, but in CanReg4 there are several columns for several sources. The tool will ask for one source but you can add as many sources as you want.

In Córdoba’s Cancer Registry there were 2 sources for each patient.

After you complete all the variables for sources, click on the “Finish” button.



1. The tool will ask for a folder where the 3 datasets (patient.csv, tumour.csv and source.csv) are going to be saved.



Note: the tool checks that the number of patients and tumours is correct.

If you go to the folder you chose you will find: patient.csv, tumour.csv and source.csv. This 3 files are the ones you have to import into your CanReg5 database.

