CanReg 5 Future Improvements

By: Morten Johannes Ervik, CSU, IARC

October 14, 2020

Introduction

This document is a road map for the development of CanReg5.

CanReg 5.1

Remaining ideas/fixes for CanReg 5.1

Data Entry related

- Design a way to streamline the export and import functionality when centralizing data.

  – Using the bulk import and export procedures from the database? (http://docs.oracle.com/javadb/10.4.1.3/tools/ctoolsimport16245.html)

Database related

- Move (parts of) the XML to the database so that we no longer have mismatch.

  – During backup we export it.
  – During start up of server we first get the XML from the server.
  – Stored as a BLOB?
  – Check compatibility of XML?
    * ruby code:

      ```
      h = {}; s.split(",").each {|b| a = b.split(" "); h[a[0]]=a[1..-1].join(" ")}
      ```
    * and compare the map/hash of the server and the file

- Better database connection security.

- Move deprecated/obsolete cases to a different table?

- A log of changes table monitoring any change to any element in the database...

- Letting users add variables (tables) that are linked to certain (ranges of) variables. (For example cancer sites, breast woudl show certain fields only relevant to breast cases...)

- Optimize queries.

Analysis related

- Implement all the analysis tables proposed.

  – Number of sources per record. (See Chapter 9 of SOP Manual (Parkin et al))

- Others:
  * Table of rates by year with the same groups as the curernt incidence table? (DF)

- Group by other meta variables for frequencies by year (age group, 2 digit icd10 etc)

- Export report automatically opened in the System? [DONE]

- An intermediate question asking for sub pop when automatically picking population data sets?

Quality control

- First name sex improved

- Improve person search

  - Look into using the Duke library!
  - Improve the dedicated merge patient interface.
    * A table is displayed, where the top line contains the original record you are editing, and other lines contain possible matches. You must tick the corresponding box on the left if you are sure it is the same person. (Tick the appropriate "Display variables" boxes to adjust what variables are displayed.) If you tick a record, then you must perform the "Multiple Primary Test" to see whether these two tumours are similar (and therefore duplicates) or not (multiple primaries). In case of duplicates, you must make one good record with all new information updated, and delete the other. In the case of Multiples, both records are kept, and the MP sequence and total boxes are automatically filled in.
    * Highlight differences between records.
  - Accept all (without looking) after a search run?
  - Keep a table of all confirmed false positives?

- Local edits

  - xml (variables needed, values accepted: ... , cominations accepted:...)

- Standard view of completeness of a certain variable. (Fex: 42% of cases has stage.)

- Missing values % (per year)

- Optimize code.

Management related

- Database structure editor:

  - Better detection of missing mandatory variables
    * Database workings
      · Source ID etc
    * Registry workings
      · Topo, Morpho, Address etc.
  - Variables grouped by table/on screen group
  - Detect duplicate variable names

- Possibility to change the path of the database/other system variables.

Pending bugs

- Merge issues

(Computer) Code related improvements

- More unit tests.

Coding related improvements

- Inconsistencies IARCcrgTools vs CanReg in ICD-O-3 to 10 conversions. (Korea)

Proposals for CanReg 5.2

Ideas/fixes for CanReg 5.2.

General

- Internalize some dictionaries (this has started, but can be taken further)

  - ICD-O-3 (Topo, Morpho, Behaviour)
    * Specialized fields?
  - Record Status.
  - Check Status.
  - Person Search Status.

- The possibility to select premade dictionaries during set up?

- Migrate GUI away from Swing Application Framework.

  - Migrate towards other GUI designers?
    * JFormDesigner? (http://www.formdev.com/jformdesigner/)
  - Migrate to browser based GUI?

Data Entry related

- Let user enter dates in the format they wish? [DONE]

- Support for some other date formats?

  - Buddhist
  - Other? Chinese?
  - Store those as additional fields and

- Improve workflow during data entry.

  - Use another GUI/form system? http://www.miglayout.com/ ?

- Build a dictionary chooser into the record editor?

- Import excel files using http://jexcelapi.sourceforge.net/

- Import from other sources improved - "passive data collection"

  - "holding table"?
  - automatic linkage proposals with patients already in the database

- Data entry using smart phones/tablets?

  - Proof of concept ready using open source tools

- Browser based data entry?

Database related

- Mortality data?

Analysis related

- Use of mortality data in analysis?
- Survival?
  - Link to the UN life tables for relative survival?
- Export excel files
  - using http://jexcelapi.sourceforge.net/
- GUI for redefining groups etc of tables?
- Export to EpiInfo?
  - What version?
  - Format?
- Export as HTML+JS with help of graph libraries?
  - http://d3js.org
  - http://cyberpython.github.com/AwesomeChartJS/
  - http://code.google.com/p/flot/
- Launch a web server and serve a live view of the databaset?
  - Using Spark? http://www.sparkjava.com/index.html
- Capture/recapture analysis

Management related

- Changing some user rights levels
  - Hiding the browser for a user-level between analyst and registrar?
- Adding dedicated procedures to deal with central databases.
- Streamline the setup of new systems
  - New system wizard
- Implement other algorithms for names sounding alike during person search
  - Double Metaphone (http://commons.apache.org/codec/apidocs/org/apache/commons/codec/langu
    * Limitations: only A-Z