



Anonymizer – Batch processing

It anonymizes the set {CTset, StructureSet, Plan, Dose} preserving the logic relationship between objects.

The anonymization is done using the MATLAB function *dicomanon* (rev R2010a) which anonymizes as per stipulated on [PS 3.15-2008 Table E.1-1](#) (*). We are not able to test full conformity of this MATLAB tool to DICOM de-identification standards, thus we do not take any responsibility for the use of this tool.

For keeping track of some possible useful details, the Patient Age is preserved declaring a fake Birth Date. Patient sex is also preserved.

Introduction:

The logic relationship between objects {CTset, StructureSet, Plan, Dose} is done in 2 main ways: Each object makes reference to its 'parent' object through the parent's UID, for instance each contour defined on the structure set is linked to the CT image where this contour is defined referencing the CT UID. On top of that, all the objects must share the same frame of reference and belong to the same study, DICOM objects whose UID normally doesn't get changed during anonymisation.

This software uses the following MATLAB functions for de-identifying and building the new logically-connected objects:

dicomanon: this function creates new dicom objects anonymizing them as per previous reference (*)

dicomid: this function creates new UIDs and is used for generating the new UIDs that does not get touched by the *dicomanon*, as the UID of the new shared frame of reference, study UID, etc.

dicomwrite: this function saves new dicom objects assigning new instance UIDs each time is invoked, this particular behaviour produce the difficulty that for knowing the UID of an object it has to be previously saved. This make the software quite slow, as many write and read cycles are required. Anonymising a H&N patient with ~20 structures requires ~6minutes.

dicomread, *dicominfo*: this pair of functions allows to read the image (read) and the metadata (info) of nearly any dicom object. I have no encountered yet a situation where they fail but this can eventually happen.

The software is not at professional grade, thus is not foolproof: users must be careful on building the input in the specified way: for instance, there is no previous check to insure the input set {CT-STRCT-PLAN-DOSE} is actually a logically connected set, thus if you provide CTs and STRT from one patient and plan from another and yet dose from another, the software will still produce at the end a logically connected set... thus, scrap-in means scrap-out!

The PLAN objects get their beams defined on a machine called "VicRapidPlan" which is a copy of the Varian testing machine "EclipseCAP_TB", a TrueBeam machine that contains many different beam types. That was intended to allow centers to import beams that they do not actually have defined on their machines, as on top of the EXTERNAL ID CONVERSION you need to have the same kind of beam: FFF or SRS beams would fail to load on machines not having this kind of beams defined. Very likely most of the centers have already this CAP machine declared into their ARIA systems, if not, the XML descriptor of this machine is provided on the github site (VicRapidPlan_TB.xml). You can import this machine for easing the managing of plans with beam types not matching the ones declared on your system



How to install:

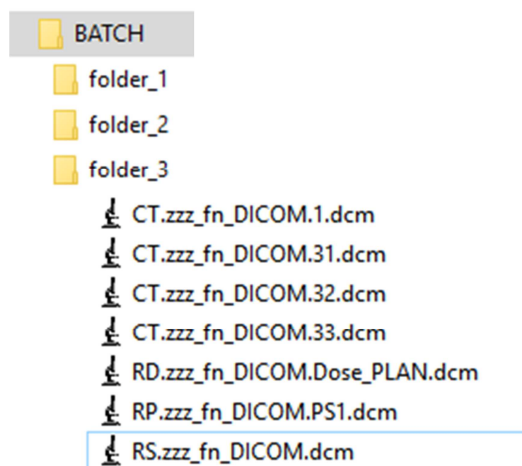
Prerequisite: you need to have [MATLAB Runtime environment](#) version 9.0.1 (64 bits) or newer installed to run the exe file. Alternatively, If you have MATLAB you can [ask](#) for the source files for running interactively (and/or to custom them for your needs).

Clone (or download) from [Github](#) the full folder 'for_redistribution_files_only'. All the files must be on a single folder (you can rename the folder as you want)

How to use:

Create a "root" folder on which each set {CTset, StructureSET, Plan, Dose} will be stored on a different subfolder of that root folder.

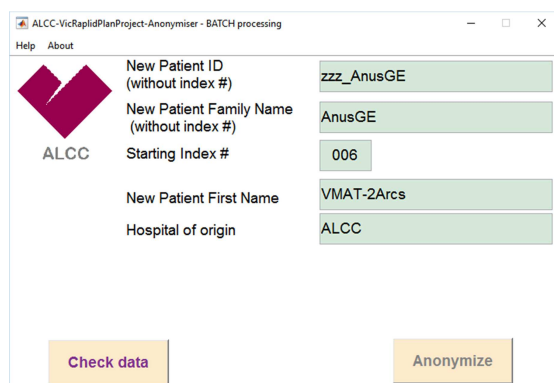
Export each set {CTset, StructureSET, Plan, Dose} to a single sub-folder. Only 1 plan, 1 structure set and 1 Dose are allowed to be on the folder with their corresponding the CT image set.



The software does not check for consistency of this set, thus it is your responsibility to export a consistent set, i.e. the structure set defined on the CT set, the plan defined on the structure set and the dose corresponding to this plan.

Run the exe file 'ALCC_AnonymizerBATCH.exe'.

Set the new generic Patient ID, Patient Family Name, Patient First Name and the name of the Hospital of origin of the set {CTset, StructureSET, Plan, Dose}.



→ zzz_AnusGE006, zzz_AnusGE007 etc.

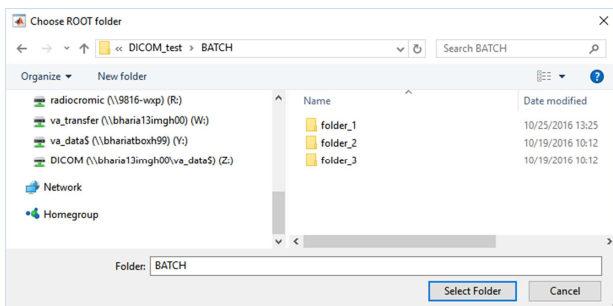
→ AnusGE006, AnusGE007 etc.



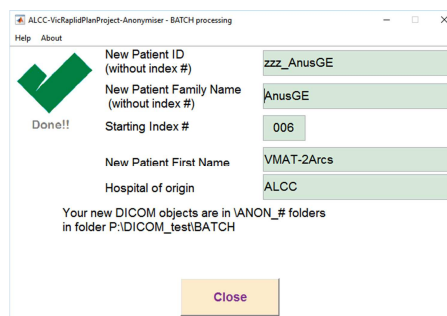
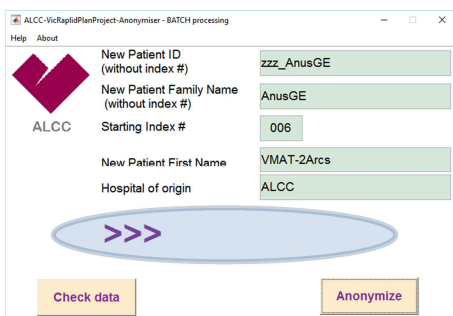
Set the Initial Index #: The PatientID and PatientLastName will have the Index # added at the end, in sequential order starting from Initial Index #: NEWID005, NEWID006, etc.

Once edition completed and checked (the software doesn't do any kind of check at this level!!), press the "Check data" to enable the "Anonymize" button that has to be pressed to launch the process.

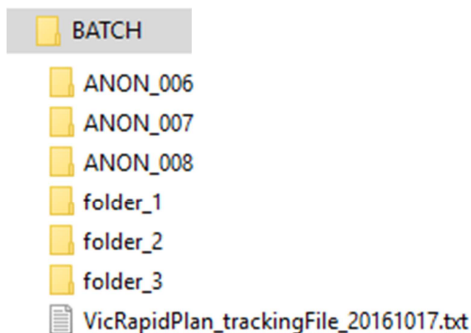
A dialog box will pop-up asking for selection of the root folder (.\BATCH on the example). A first check is done to detect any flagrant fault on the input data: it checks for having CT data (at least 1 image), and for the presence (and uniqueness) of the remaining objects: PLAN, StructureSet and DOSE.



Then the software starts to do its many cycles of read write. For having an indication that it is actually working a moving arrow is shown which moves after each read or write event: note that when writing big objects (as Doses) it gets immobile for awhile.



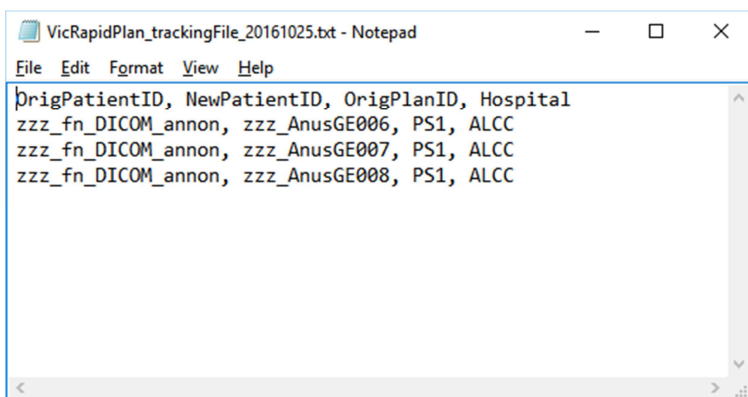
The result is a series of new folders .\BATCH\ANON_### (created on the root folder) containing the new anonymized DICOM objects.



A text file named "VicRapidPlan_trackingFile_yyyymmdd.txt" is created on the root folder containing the following information: Original PatientID, New PatientID, Original PlanID, Hospital of origin.



The first line have the headers of the data and following lines the data itself. Its purpose is to facilitate managing the plan flow into models. Be aware it contains the Original Patient ID.

A screenshot of a Notepad window titled "VicRapidPlan_trackingFile_20161025.txt - Notepad". The window contains a text file with four lines of data. The first line is a header with five comma-separated fields. The following three lines are data rows, each containing the same five fields with specific values. The text is as follows:

```
OrigPatientID, NewPatientID, OrigPlanID, Hospital  
zzz_fn_DICOM_annon, zzz_AnusGE006, PS1, ALCC  
zzz_fn_DICOM_annon, zzz_AnusGE007, PS1, ALCC  
zzz_fn_DICOM_annon, zzz_AnusGE008, PS1, ALCC
```

The original DICOM files are not changed.

There is also a version we developed for education purposes: it does the anonymization but does not change the original machines (linac, scanner), thus allowing for the reimport of the anonymised set {CTset, StructureSET, Plan, Dose} on the clinical data base preserving the possibility of calculating doses with the originally defined beams. You can download this version [here](#).