Protocol: import a case into the shoulder database

This document describes the protocol to add a case (set of CT scans) into the shoulder database. After the preliminary notes below, the protocol is described in two phases: 1) Copy CT scans in pending directory, 2) Import CT scans from pending directory. Finally, an alternative protocol is described at the end of the document.

Preliminary notes

- Data
    - They are in *lbovenus.epfl.ch:/home/shoulder/data/*.
    - This directory should be mounted locally.
    - Each entry of the database is a shoulder case (***SCase***) associated to a patient.
    - Each SCase is uniquely identified by a SCaseID.
    - There are two SCase types or groups: normal (N) and pathologic (P).
    - In the pathologic group, each SCase corresponds to the treatment of the pathology of a shoulder. Since patients may have been operated on both shoulders, possibly twice on the same shoulder, the same patient might be associated to four SCases. Besides, a SCase may be associated with several CT scans, for the same patient and shoulder.
    - Patients of the normal group appear only once, but may have both shoulders in the CT scans, and by definition, have no pathology. They usually have a trauma elsewhere.
    - The import process consists in importing CT scans into a SCase, which might be created, or updated
    - CT scans are coded with DICOM (Digital Imaging and Communications in Medicine) format, which is a worldwide-accepted standard for medical images exchange, managed by the Medical Imaging & Technology Alliance – a division of the National Electrical Manufacturers Association (*https://www.dicomstandard.org/current/*).
    - After importation, CT scans are classify as follows:
        - Group of patients:
            - **Normal (N)**: CTs from patients without shoulder pathology. Those CT scans are often acquired in order to find multiple traumas after an accident, so they might contain a much larger portion of the patient body than the area of interest of this database.
            - **Pathological (P)**: CTs from patient with shoulder pathology. Those CT scans are aimed to analyze the shoulder anatomy and can contain several CT sets with different acquisition parameters.
        - Acquisition stage:
            - **Preop** CT sets acquired before the shoulder joint replacement surgery either for diagnostic or surgery planning purposes. Every CT set without prostheses in the shoulder of interest must be classified as *preop*.
            - **Postop** The CT sets acquired after the surgery aiming to validate the prostheses placement (just after surgery) or checking their state (some time after the surgery). This means that every CT with a shoulder prostheses on the shoulder of interest should be initially considered as *postop* (unless the responsible clinician informs explicitly about the eventual case of a prostheses replacement).
        - Body part:

- **Shoulder** is the tag for every CT containing (at least) the shoulder of interest.
- **Elbow** is the tag for CTs containing elbow without shoulder.
- **Contralateral** or *Opposed Shoulder* is the tag for CTs with the opposed shoulder to the shoulder of interest.
- **Thorax** is the tag for the CTs containing the patient thorax without any complete shoulder.
- **Other** is the tag for any other body.

- Phantoms (**with** or **without**):
  - These are samples of synthetic bone with accurate and homogeneous mineral density and cylindrical shape. Phantoms are typically placed behind the back of the patient and are used to calibrate the measurement of the bone mineral density. At least 2 phantoms are needed for an accurate calibration.
  - The following table summarises CT classification adding a **CT series label** as a numerical value for each enumerated class (column "#")

Acquisition	Body	Phantoms	Kernel	#
Preop	Shoulder	Without	Sharp	1
Preop	Shoulder	Without	Smooth	2
Preop	Shoulder	With	Sharp	3
Preop	Shoulder	With	Smooth	4
Preop	Elbow	-	Sharp	5
Preop	Elbow	-	Smooth	6
Preop	Thorax	-	Sharp	7
Preop	Thorax	-	Smooth	8
Preop	Opposed Shoulder	-	Sharp	9
Preop	Opposed Shoulder	-	Smooth	10
Preop	Other	-		11
-	-	-	-	-
Postop	Shoulder	Without	Sharp	p1
Postop	Shoulder	Without	Smooth	p2
Postop	Shoulder	With	Sharp	p3
Postop	Shoulder	With	Smooth	p4
Postop	Elbow	-	Sharp	p5
Postop	Elbow	-	Smooth	p6
Postop	Thorax	-	Sharp	p7
Postop	Thorax	-	Smooth	p8
Postop	Opposed Shoulder	-	Sharp	p9
Postop	Opposed Shoulder	-	Smooth	p10
Postop	Other	-		p11

- Scripts
  - Can be executed locally

- Require the graphical interface
- They are in *lbovenus.epfl.ch:/home/shoulder/methods/matlab/database/*
- They can be pulled from remote git repository
  https://c4science.ch/diffusion/8218/shoulderdb_importsegmentmeasureanalyze.git
- The import script is located in the sub-directory *Import* (to be renamed ImportSCase), and should be executed from there.
- Each patient folder pending to be imported should be initially considered as a different SCase. When more than one set of CTs are present for a single SCase, we might find different types of CT images: scout-scans, shoulder CTs, elbow CTs, thorax CTs, opposed shoulder CTs, etc. Not all of them must be imported, however this will be explained later in this document (more information in section *CT types classification*).
- The anonymization is performed during importation (to be modified before 2020). This consists in removing some information from the metadata of the DICOM files that can identify the patient, such as its name or birth date. This can be automatically done by the Matlab script or manually with Amira.
- After importation, the script *dicominfoSCase.m* should be executed to update the Excel files CTdicomInfo.csv and CTdicomInfo.xls, located in ***shoulder/method/database/***. File CTdicomInfo.xls is linked to the main Excel file ***ShoulderDataBase.xlsx***. (this is now executed when bottom **"Add new SCase to Excell Database"** is pushed)

Protocol

1. Copy CT scans in pending directory

1. Extract the archive (.zip) file sent by CHUV. Inside the unzipped folder, each patient folder, is named with the *Patient Permanent Identifier*, **IPP** (*Identifiant Permanent du Patient*), and contains at least one sub-folder with a CT reconstruction set of DICOM images.These DICOM files are CT scans that we will be imported.
2. For normal SCase, save the unzipped folder in the pending directory:
   - */data/N/pending/N-update-[YYYY-MM-DD]*, where [YYYY-MM-DD] is the date you received the archive
3. For a pathologic SCase, save the unzipped folder in the pending directory: */data/P/pending/TSA-update-[YYYY-MM-DD]*, where [YYYY-MM-DD] is the date you received the archive

   Example: */data/P/pending/TSA-update-2018-07-26*

2. Import CT scan from pending directory

In this subsection, we will refer to the numbered paragraphs as *step1*, *step2*, etc.

1. Run the Matlab function ***importSCase.m*** located in *shoulder/methods/matlab/database/Import* (**this directory will change to *shoulder/method/database/ImportSCases***)
   - A prompted window will ask you to select the *data* directory, which contains all imported SCases and the Excel database.

2. Select **Normal** or **Pathologic** in *Patient Group*, depending on the type of CT that you are about to import.
   - If you choose *Pathologic*, a message window will be prompted, asking you whether to activate or not the 3D visualisation. We recommend to avoid the 3D viewer activation at this step, because it may be activated later. For more information read the *step5* of this section.

3. If you are importing a new patient set of CTs !!AT: not clear!! (first CT of a new pending SCase), click on button ***Start new Patient/SCase*** and select the patient folder of your interest. Otherwise, you might

skip this step.

- IMPORTANT: select the folder that contains the sub-folders with the CTs. For example:
    - The pending patient directory *lbovenus.epfl.ch:/shoulder/data/P/pending/TSA-update-2017-08-18/97212* contains:
        - *DICOMDIR* (file)
        - *LOCKFILE* (file)
        - *VERSION* (file)
        - *24NZUR5U* **(Folder to be selected in this step)**
            - This folder contains the *sub-folders CO1ALS1N*, *CO02LW5N*, *FFNFXVOY*, *FMNFXVOY* ans *RMEJ0LXY*, each one of which contains a CT series.
- Information: This step...
    - ... finds the number of slices present in each of the CTs of the new patient.
    - ... adds a prefix to the CTs sub-folders with the number of slices contained so they can be prioritized depending in the CT size.
    - ... adds a suffix that help discriminate between potential shoulder CTs, elbow CTs or scout CTs.

4. Click on ***Load next CT set*** to load a new CT set.
    - Navigate inside the patient folder until finding the set of sub-folders that contain the CTs. These sub-folders have now prefixes indicating the number of CT slices inside. Start loading the largest CT (usually *shoulders* are in the largest CTs) and continue in decreasing order.
    - Select all DICOM files (not file "VERSION") in the folder and click on button *"Open"*.
        - A bar-progress window will be prompted showing the loading progress. Don't close it before the process is finished, because this will stop the loading and you will need to start again this *step4*.
        - The CT slices will appear in main viewer (2D) and (if *3D visualization* is activated) a 3D rendering will be shown in the 3D viewer as well.
        - The information box above the 2D viewer shows the full-name of the current slice.
    - During loading, some types of errors might appear:
        - Window prompted with *Error identifier:* **heterogenousStructAssignment**:
            - This means that at least one of the selected slices has a different format, so they cannot be loaded as a single CT set.
            - Try to identify the problematic images and then load again the CT set but skipping those files:
                - Start by checking in which image the loading progress bar has stopped, because this must be the file producing the error.
                - Load the rest of the DICOM images by selecting all of them except the problematic one.
                - If the problem happens again it means that there are several DICOM files with different format. Hence, repeat the process a few times to avoid all the problematic files.
            - Alternatively, open the full CT set with another software like Amira, which is able to differentiate the problematic slices. Then, saving the set of correct images in DICOM format

within a different folder should allow further loading of the new set without the *heterogeneous error*.

- This process might divide the CT set in two or more parts (when the problematic file is neither the first nor the last).

- If the error is still present it might be necessary to use Amira to import this CT set, but it could be also due to an error made in the reconstruction step.

  - Check if it opens correctly in Amira but if it creates several series of images and you cannot identify any of them as relevant it might be better to avoid the CT set. Eventually, shall the CT provider be informed.

- Do not import unusual CT image formats (atypical orientation, slices from a 3D rendering, images with superposed grid or measurements, etc).

  - Some of these unusual formats are treated by the script in such a way that the containing folder will be automatically renamed adding the suffix *wrongOrientation*. The user just needs to avoid these sets and continue loading the next one.

- If the Matlab script cannot read the DICOM file metadata

  - The alternative is to import CT with **Amira (described later in this document)**.

5. (optional) If needed, activate the 3D viewer.
   - The 3D visualization is useful for shoulder side identification of Pathological series, but rarely needed in Normal patient because Normal CTs usually contain both shoulders.
   - WARNING: when the CT are very large (over 400 slices) and the computer has a limited memory the loading might slow down or even stop the script.
   - When activated it shows a *volume rendering* of the CTs.

6. Select **Left** or **Right** in the radio-buttons group *Shoulder Side*.
   - The side selected must match with the *shoulder of interest* and not with the *contralateral shoulder* if this is present. (For more information read *Preliminary notes* and *CT types classification* above in the current document)
   - For Normal CTs were both shoulders are present in resting position, select *Both* shoulders (if the importation software doesn't have this option choose *Right* shoulder).

7. Select **Preop** or **Postop** in the radio-buttons group *Acquisition stage*

8. Click in button **Auto Check DB** to launch the database auto-check for IPP coincidences.
   - A window will be prompted asking you to select the Excel file containing the database. This is *ShoulderDataBase.xlsx*, located in *lbovenus.epfl.ch\data\Excel* (or *lbovenus.epfl.ch\data\Excel* for development purposes).
   - The results will be shown in the *DB check result* information box.
   - Depending on the auto-check results, a *SCaseID* will be suggested. But if any problem rises, you might always do a manual selection of the *SCaseID*, as described in the next *step9*.

9. (optional) If the *Suggested SCaseID* does not match with your expectations, you might choose manually the output directory by clicking in button **Manual SCase**.
   - Example: *shoulder/data/P/5/5*

10. Verify the *Parent output directory* was updated correctly or modify it manually if needed.

- Example: *lbovenus.epfl.ch:/shoulder/data/P/5/5*

11. Choose between the different possibilities in the radio-buttons group *Body part*:

    - **Shoulder** must be chosen when at least one complete shoulder is present in the CT reconstruction.
    - **Elbow** must be chosen when the CT contains at least one elbow without any complete shoulder.
    - **Thorax** must be selected only when the main body in the CT is the *thorax* and no shoulder is present or at least not complete.
    - **Contralateral** must be selected when the only complete shoulder present in the CT is not the *shoulder of interest*.

12. In *CT Kernel (image filter)*, select **Sharp (bone)** when the 2D viewer shows a *sand like* texture for the CT slices or **Smooth (soft tissue)** the image looks soft and smooth.

13. Select **With** or **Without** in the radio-buttons group *Phantoms*.

14. (optional) When a SCase contains many CT sets, it might be very useful adding some extra information to the *README* file, so the CTs can be differentiated easily. As examples:
    - When both shoulders are present and complete --> add *both shoulders*
    - When several CT series can be classified similarly in terms of body part or CT Kernel, it might be useful adding other CT data.
    - This is also the place where the operator might add any comments

15. Check in the readme line below the 2D viewer that all the options selected are correct and click on button **Import to Database**.
    - If the imported CT is meant to be the first CT of a new SCase in the database, this will create a new folder in the directory indicated in the text field *Parent output directory* with the SCase identification number indicated in *Suggested SCaseID* and the IPP found in the metadata (and shown in the field *IPP*, in the *DICOM Metadata* box). Example: *shoulder/data/P/5/5/P555-918819*.
    - Inside the SCase folder, a new sub-folder will be created to store the imported CT images:
        - The name of this sub-folder contains the SCaseID, the IPP and the CT set numeric label. Example: ***CT-[SCaseID]-[IPP]-#*** --> ***CT-P555-229341-1***
            - A new line will be added to the *README* line with the name of the new folder followed by some information like the *kernel*, the presence of phantoms, the main body in the CT, etc.
            - If the CT set label already exists in the SCase folder, then a new folder will be automatically created with *"xxx"* as CT set label. Consequently, the CT series label will need to be updated manually. And the corresponding line in the *README* file will need to be corrected.
        - The anonymized DICOM images will be finally stored inside the directory like ***shoulder/data/P/5/5/P555-918819/CT-P555-229341-1/dicom***.

16. If the recently imported CT set was a new SCase, the Excel Database need to be updated. In that case, click on button *Add new SCase to the Excel DB*. This automatically updates the Excel Database with almost all the DICOM information from the imported SCase. However, this parameter (SCaseID) have to be manually added to the Excel Database:
    - Open the file ***ShoulderDataBase.xlsx***, located in *//lbovenus.epfl.ch/shoulder/data/Excel/*.
    - the SCaseID of the last entry in the database for your case type (N or P) and increment by 1 to get the current SCaseID (g.e. if the last entry in database is *P284*, current

SCaseID will be *P285*").

- Scroll down to find the last SCaseID (last row) and add a new row by taping the new SCaseID in the cell below the last SCaseID (first column). This automatically add a new row to the table, and copy formulas. Then:
- In the second column introduce the row number by simply increasing the number from the immediately superior row.
- From the *DICOM Metadata box* check that the following data are correctly in their corresponding columns of the excel file **ShoulderDataBase.xlsx**:
  - IPP
  - Patient Birthday
  - Patient Gender
  - Patient Initials
- Check also the shoulder side column *shoulder_side*
- Check that the rest of the columns are filled with data as in the precedents rows. If you find some columns with *NaN* values:
- close the Excel File without saving,
- try to run manually the script *dicominfoSCase.m* without arguments, which is located in *shoulder/methods/matlab/database*. This script updates the new row in the Database with metadata from the imported CT.
- try to solve the errors if some arise
- open the Excel file and start this *step16* again.
- Save and close the Excel file.
- Run the script **dicominfoSCase.m**, which is located in *shoulder/methods/matlab/database*. This script updates the new row in the Database with metadata from the imported CT. Once the script has finished open the Excel Database again and save it.

17. If the patient pending folder contains more CT sets to import, continue from *step4*. Otherwise, start a new patient folder and continue from *step3*.

Alternative import protocol using AMIRA

!! Might be simplified by just saving in a correct format, usable by matlab !! --> already included above ?

- When the CT has a non-uniform resolution, the (above) Matlab protocol may not work, but can be replaced by the following Amira protocol.

1. Start Amira and click on *Open Data…* (Ctrl + O). Navigate to *shoulder/data/[X]/pending/[IPP]*. Inside the CT folder, there might be several reconstruction folders.

2. Load the images of the CT folder that you could not import with the Matlab script
   - Select all files except the file named *VERSION* (if it exists).

3. Check the parameter *series description*, in the prompted window, to know the kernel used for this set in *series description*. Usually it is *OS* for **Sharp** and *STD* for **Smooth**. Click on OK.

4. Attach an *Ortho Slice* or *Slice* display object to the newly added green data modules to visualize the content of the image stacks.

5. Open Excel shoulder database *shoulder/data/Excel/ShoulderDatabase.xlsx* database (*SCase* tab) and compare the IPP, CT date, shoulder side and DICOM information to determine if the new case is already in the database.
   - If the IPP already exits in the database:

- Find the *Readme* file in the SCase folder of the existing SCase and compare the data inside with the metadata of the current CT to decide whether the current CT set already exist in the SCase folder or it have to be imported.
- If it does not exist continue with step6.

6. Right-click on the first data module and click on *Export Data As…* and choose type *DICOM*.

8. Go to the correct folder *shoulder/data/[X]/[#1]/[#2]/[CaseID]-[IPP]* , where *#1* and *#2* are the first two digits following 'P' or 'N' in the current SCaseID. For example: *shoulder/data/P/0/1* for SCase *P13*. If folder [CaseID]-[IPP] does not exist yet, create it. Then create the folder *CT-[CaseID]-[IPP]–#* and inside this folder, create a folder *dicom*.

9. Save the series with the name *[CaseID]-[IPP]_[PatientInitials]-[bone/smooth].0*, e.g. *P123-456789_XY.0*.

10. In the next window, replace the patient's name by *[CaseID]_[IPP]–[PatientInitials]*.

11. If the error: *Non-uniform coordinates not supported* appears, you will need to convert your data to uniform coordinates before exporting to DICOM.
    - To do the conversion, attach an *Arithmetic* module to the data object. Select *regular* for the result type.
    - Specify the resolution (ex: 512x512x267) similar to the original data.
    - In the *Expression field* put A (the input data object).
    - Click on *Apply*.
    - The result is a new uniform coordinate data object.
    - Save the result as described in points 7 to 10.

12. Rename the folder *CT-###-IPP-xxx* following the classification presented at the beginning of this document.

13. This step is similar to the step16 of the Matlab SOP, described in the previous section *2. Matlab interface for SCase anonymization and importation*

14. Continue with the next CT set.