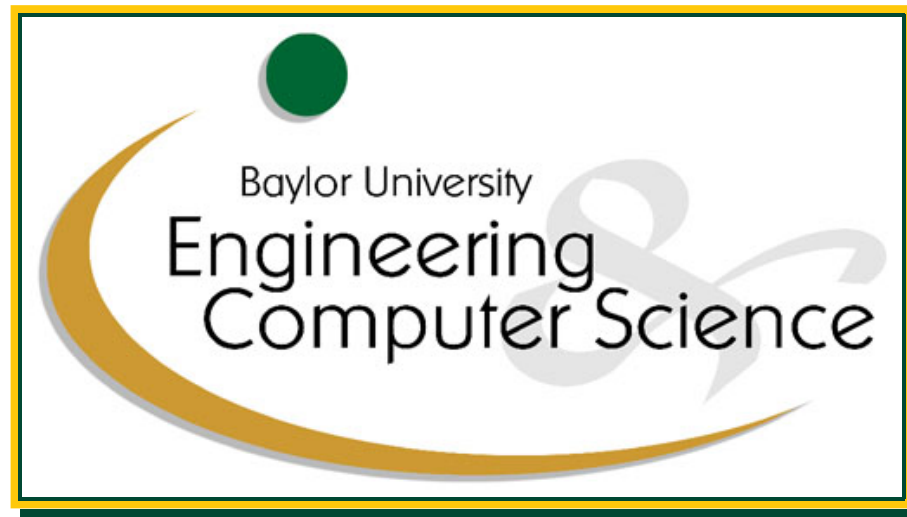# pCT Data Storage on Baylor Server

Folder/File Naming and Organizational Scheme

Blake Edward Schultze

February 7, 2016

/home/<*user name*> : On each Kodiak/Tardis node, this is the local home directory of each user where the files controlling user settings/history are stored. Although the user account settings files (.bash_profile, .bash_history, etc.) are automatically distributed to each node so that any preference changes will apply to user sessions on any node, this data is not backed up and there is a limited storage capacity. Thus, this directory should only be used to store small amounts of temporary data. These directories are private and only accessible by that user.

/data/<*user name*> : This is the global home directory of each user where the input/output data for code/program execution and all permanent/important personal data should be stored. As a subdirectory of **/data**, the contents of these directories are backed up periodically, but like the local home directories, each user only has access to their personal directory.

/ion/. . . : This directory is dedicated to proton and ion therapy research and contains directories where users can upload data and/or code and selectively share these files with the collaboration by moving the desired files to the staging area (**/staging/<*user name*>**) from which an administrator will move these to the appropriate data/code directory. The data in this directory is located on a network storage device and this drive is mounted on both the Kodiak and Tardis clusters, making it accessible from all of the master/compute nodes. The contents of this directory are periodically backed up to tape drive from which data/code can be recovered in case of failure.

/ion/incoming/<*user name*> : This is the directory where users should upload data/code to the Baylor server prior to preparing and organizing it for sharing with the collaboration, with each user only having access to their personal directory in **/incoming**. Data/code that a user wants to share with the collaboration should then be organized appropriately and then moved to their personal directory in the staging area (**/ion/staging/<*user name*>**) from which data/code is moved to a shared directory accessible by the entire collaboration. In general, users should consider removing data from their incoming directory once it has been added to the staging area and moved to a shared directory to avoid unnecessary duplicate data and memory consumption.

/ion/staging/<*user name*> : This is the directory where users should move data/code when it is ready to be shared with the collaboration and contact. At the moment, users should then notify a cluster administrator (e.g. Brian Sitton, Keith Schubert) when there is data/code ready to be moved to a shared directory, but work will be done to make this a more streamlined process. Note that administrators typically will not be familiar with pCT or the organizational scheme so it important for the files in the staging area to be organized such that their path below **/ion/staging/<*user name*>** exactly mirrors the path below **/ion** where it is to be moved.

In other words, the data in the staging directory should be organized into the exact same hierarchy of directories as it is to be organized in the target subdirectory of **/ion**, as defined by the pCT data/code organizational scheme, with the top directory in the staging area corresponding to the first subdirectory of **/ion** along the target path (often **/pCT_data** or **/pCT_code**). The administrator can then move the data/code to the appropriate

target destination by following the target and staging area paths downward from **/ion** and **staging/<*user name*>** , respectively, moving data from the staging area to the corresponding target directory as it is encountered. This is repeated until all files have been moved or a subdirectory along the target path does not exist in the target hierarchy, at which point the missing subdirectory and its contents can then be moved directly from the staging area to the target directory to complete the move.

/ion/pCT_data/... : This directory is where the raw, preprocessed, projection, and reconstruction data/images are moved to make them available to the other pCT users. Each type of data is stored in separate subdirectories and soft links to this data are created and organized in a directory hierarchy indicating their input/output data dependencies. The directory/file naming and organizational scheme for each type of data and the soft links are outlined in the next section of this document. Data/images should only be moved to this shared directory after having been verified as valid/accurate and having been organized appropriately.

/ion/pCT_code/... : This directory contains the code/programs used to generate simulated pCT data, preprocess raw experimental data, reconstruct images from experimental/simulated data, and tools for analyzing reconstructed images. This directory also contains the code/program which calculates the water equivalent depth (WED) for a given list of beam angles and beam aim point (BAP) coordinates using the reconstructed RSP values for the target volume. Each type of program and the corresponding code are stored in separate directories, and for each of these directories there is a "*master*" directory containing the code/program that should be used/executed by users. Each of these directories also contains separate subdirectories for each user involved in the development of the corresponding program and as a subdirectory of **.../ion**, this code is then accessible by other users involved in the development of this program, making collaboration with them easier. Ideally, everyone will manage their code/software development with something like GitHub in which case they could develop code from their personal directory and make their code available to collaborators through GitHub, but this is not a strict requirement (just a strong suggestion). In cases where there are multiple versions of a program being developed completely separately (as is the case with the preprocessing code and the old/new reconstruction code), there should be a "*master*" version for each of these in separate subdirectories with names making these discernable.

/ion/pCT_data/pCT_Documentation/... : Documentation relevant to pCT is stored in this directory, such as descriptions of the data format, coordinate system, and phantoms and pCT related publications (including student theses/dissertations). This is a GitHub managed local repository allowing everyone to "push" contributions to the repository and "pull" updates/additions from others into their own local clone ensuring everyone has access to the latest information.

Key:

Green : directories whose names do not change

Brown : directories whose names vary depending on object name or date

Italic/Royal Blue : individual files

Italic/Dark Blue : multiple files

/**data**/**ion**/**pCT_data**/**organized_data** : directory containing all of the pCT data (raw, processed, projection, and reconstruction), primarily as soft links to the actual data itself (with each type of data stored/organized in separate directories), organized in a hierarchy of subdirectories indicating data dependencies.

(1) /**Phantom** : directory containing all of the experimental/simulated data and reconstructed images associated with the phantom/object named "*Phantom*".

    (a) /**Reference_Images** : directory containing reference images (xCT, RSP, etc) relevant to analysis/comparison of the data/images for this object and data type.

    (b) /**Experimental** : directory containing data and images generated from an experimental scan of the object.

        (i) /**YY-MM-DD** : directory containing data and reconstructed images corresponding to all experimental scans of the object performed on this date.

            (1) /**XXXX[_AAA]** : directory containing data and reconstructed images corresponding to the experimental scan of the object for this particular run # of this date, where the run # is of the form "*XXXX*" with an optional descriptor tag "*_AAA*" added specifying additional pertinent information about the scan, such as inferior "*_Inf*" or superior "*_Sup*" positioning as in the case of a head phantom.

                (a) /**Input** : directory containing raw data generated by scan of object from each gantry angle and transmitted by event builder.

                    (i) *raw_xxx.bin* : binary files containing trigger/tracker/energy detector data from event builder associated with gantry position "*xxx*" = {"*001*", "*002*", "*003*", · · · }.

                (b) /**Output** : directory containing calibration and post processed data generated from analysis of raw data and used as input to image reconstruction.

                    (i) /**YY-MM-DD** : directory containing the post processed "*projection_xxx.bin*" data generated on this date and the reconstructions using this data.

                        (1) *calib.txt* : text file containing calibration curve coefficients for WEPL calibration.

                        (2) *TVcalib.txt* : text file containing calibration curve coefficients for tv corrected WEPL calibration.

                        (3) *settings.cfg* : specifies scan properties such as gantry angle interval, t/v detector size, reconstruction volume dimensions, etc and initial settings to use for preprocessing/reconstruction.

                        (4) *projection_xxx.bin* : binary files containing tracker coordinates and WEPL data associated with gantry position "*xxx*" = {"*001*", "*002*", "*003*", · · · } converted from raw data using tracker alignment, track reconstruction, and WEPL calibration routines and used as input to image reconstruction.

(5) **/Reconstruction** : directory containing preprocessed data and reconstructed images generated using the "*projection_xxx.bin*" data along with reference images relevant to the object.

    (a) **/YY-MM-DD** : directory containing the preprocessed data generated on this date and the reconstructed images generated from this data.

        (i) *settings_log.cfg* : copy of **settings.cfg** with any changes made to parameters/options applied at execution, if any.

        (ii) *execution_times.txt* : execution times for various portions of preprocessing and/or reconstruction and total program execution time.

        (iii) *bin_counts.txt* : linearized bin # for each proton history, where linearized bin # = t_bin + angle_bin * T_BINS + v_bin * T_BINS * ANGULAR_BINS.

        (iv) *mean_rel_ut_angle.txt* : mean relative ut angle ($\angle ut_{out} - \angle ut_{in}$) by linearized bin #.

        (v) *mean_rel_uv_angle.txt* : mean relative uv angle ($\angle uv_{out} - \angle uv_{in}$) by linearized bin #.

        (vi) *mean_WEPL.txt* : mean WEPL value by linearized bin #.

        (vii) *stddev_rel_ut_angle.txt* : standard deviation of the relative ut angle ($\angle ut_{out} - \angle ut_{in}$) by linearized bin #.

        (viii) *stddev_rel_uv_angle.txt* : standard deviation of the relative uv angle ($\angle uv_{out} - \angle uv_{in}$) by linearized bin #.

        (ix) *stddev_WEPL.txt* : standard deviation of the WEPL value by linearized bin #.

        (x) *sinogram.txt* : mean WEPL after statistical cuts with the t_{bin} and angular bin θ_{bin} plane for each vertical bin v_{bin} stacked on each other.

        (xi) *hull.txt* : text file specifying hull in 1s/0s with the xy plane for each slice stacked on each other.

        (xii) *FBP.txt* : text file specifying filtered back projection image with the xy plane for each slice stacked on each other.

        (xiii) *x_0.txt* : text file specifying voxel values of initial iterate with the xy plane for each slice stacked on each other.

        (xiv) *sin_table.bin* : file containing the tabulated values of sine function

        (xv) *cos_table.bin* : file containing the tabulated values of cosine function

        (xvi) *coefficient.bin* : file containing the tabulated scattering coefficient values for Σ_1/Σ_2 for $u_2 - u_1/u_1$ values

        (xvii) *poly_1_2.bin* : file containing the tabulated MLP polynomial values with coefficients $\{1, 2, 3, 4, 5, 6\}$

        (xviii) *poly_2_3.bin* : file containing the tabulated MLP polynomial values with coefficients $\{2, 3, 4, 5, 6, 7\}$

        (xix) *poly_3_4.bin* : file containing the tabulated MLP polynomial values with coefficients $\{3, 4, 5, 6, 7, 8\}$

        (xx) *poly_2_6.bin* : file containing the tabulated MLP polynomial values with coefficients $\{2, 6, 12, 20, 30, 42\}$

(xxi) *poly_3_12.bin* : file containing the tabulated MLP polynomial values with coefficients $\{3, 12, 30, 60, 105, 168\}$

(xxii) *MLP.bin* : binary file with MLP path data for each history entering hull.

(xxiii) *WEPL.bin* : binary file specifying WEPL value for each history entering hull.

(xxiv) *histories.bin* : binary file specifying entry/exit coordinates/angles, bin number, gantry angle, and hull entry x/y/z voxel # for each history entering hull.

(xxv) **/Images** : directory containing reconstructed images generated using this preprocessed data.

    (1) **/YY-MM-DD** : directory containing the reconstructed images generated on this date using the preprocessed data above.

        (a) *x_k.dcm* : DICOM images of x after k iterations.

        (b) *x_k.txt* : text images of x after k iterations.

        (c) *x_k.png* : PNG images of x after k iterations.

(c) **/Simulated** : directory containing data and images generated from a simulated scan of the object.

    (i) **/G_YY-MM-DD** : directory containing data and reconstructed images corresponding to all GEANT4 simulated scans of the object generated on this date.

        (1) **/XXXX[_AAA]** : directory containing data and reconstructed images corresponding to the experimental scan of the object for this particular run # of this date, where the run # is of the form "*XXXX*" with an optional descriptor tag "*_AAA*" added specifying additional pertinent information about the scan, such as inferior "*_Inf*" or superior "*_Sup*" positioning as in the case of a head phantom.

            (a) **/Input** : directory containing raw data files generated by simulated scan of object for each gantry angle.

                (i) *raw_xxx.bin* : binary files containing trigger/tracker/energy detector data from event builder associated with gantry position "*xxx*" $= \{$"*001*", "*002*", "*003*", $\cdots\}$.

            (b) **/Output** : directory containing calibration and post processed data generated from analysis of raw data and used as input to image reconstruction.

                (i) **/YY-MM-DD** : directory containing the post processed "*projection_xxx.bin*" data generated on this date and the reconstructions using this data.

                    (1) *calib.txt* : text file containing calibration curve coefficients for WEPL calibration.

                    (2) *TVcalib.txt* : text file containing calibration curve coefficients for tv corrected WEPL calibration.

                    (3) *projection_xxx.bin* : binary files containing tracker coordinates and WEPL data associated with gantry position "*xxx*" $= \{$"*001*", "*002*", "*003*", $\cdots\}$ converted from raw data using WEPL calibration routine and used as input to image reconstruction.

(4) **/Reconstruction** : directory containing preprocessed data and reconstructed images generated using the "*projection_xxx.bin*" data along with reference images relevant to the object.

    (a) **/YY-MM-DD** : directory containing the preprocessed data generated on this date and the reconstructed images generated from this data.

        (i) *hull.txt* : text file specifying hull in 1s/0s.

        (ii) *FBP.txt* : text file specifying filtered back projection image.

        (iii) *x_0.txt* : text file specifying voxel values of initial iterate.

        (iv) *MLP.bin* : binary file with MLP path data for each history entering hull.

        (v) *WEPL.bin* : binary file specifying WEPL value for each history entering hull.

        (vi) *histories.bin* : binary file specifying entry/exit coordinates/angles, bin number, gantry angle, and hull entry x/y/z voxel # for each history entering hull.

        (vii) **Images/** : directory containing reconstructed images generated using this preprocessed data.

            (1) **YY-MM-DD/** : directory containing the reconstructed images generated on this date using the preprocessed data above.

                (a) *x_k.dcm* : DICOM images of x after k iterations.

                (b) *x_k.txt* : text images of x after k iterations.

                (c) *x_k.png* : PNG images of x after k iterations.

(ii) **T_YY-MM-DD/** : directory containing data and reconstructed images corresponding to all TOPAS simulated scans of the object generated on this date.

    (1) **XXXX[_AAA]/** : directory containing data and reconstructed images corresponding to the experimental scan of the object for this particular run # of this date, where the run # is of the form "*XXXX*" with an optional descriptor tag "*_AAA*" added specifying additional pertinent information about the scan, such as inferior "*_Inf*" or superior "*_Sup*" positioning as in the case of a head phantom.

        (a) **Input/** : directory containing raw data files generated by simulated scan of object for each gantry angle.

            (i) *raw_xxx.bin* : binary files containing trigger/tracker/energy detector data from event builder associated with gantry position "*xxx*" = {"*001*", "*002*", "*003*", · · · }.

        (b) **Output/** : directory containing calibration and post processed data generated from analysis of raw data and used as input to image reconstruction.

            (i) **YY-MM-DD/** : directory containing the post processed "*projection_xxx.bin*" data generated on this date and the reconstructions using this data.

                (1) *calib.txt* : text file containing calibration curve coefficients for WEPL calibration.

                (2) *TVcalib.txt* : text file containing calibration curve coefficients for tv corrected WEPL calibration.

(3) **projection_xxx.bin** : binary files containing tracker coordinates and WEPL data associated with gantry position "*xxx*" = {"*001*", "*002*", "*003*", ⋯} converted from raw data using WEPL calibration routine and used as input to image reconstruction.

(4) **Reconstruction/** : directory containing preprocessed data and reconstructed images generated using the "*projection_xxx.bin*" data along with reference images relevant to the object.

  (a) **YY-MM-DD/** : directory containing the preprocessed data generated on this date and the reconstructed images generated from this data.

  (i) *hull.txt* : text file specifying hull in 1s/0s.

  (ii) *FBP.txt* : text file specifying filtered back projection image.

  (iii) *x_0.txt* : text file specifying voxel values of initial iterate.

  (iv) *MLP.bin* : binary file with MLP path data for each history entering hull.

  (v) *WEPL.bin* : binary file specifying WEPL value for each history entering hull.

  (vi) *histories.bin* : binary file specifying entry/exit coordinates/angles, bin number, gantry angle, and hull entry x/y/z voxel # for each history entering hull.

  (vii) **Images/** : directory containing reconstructed images generated using this preprocessed data.

  (1) **YY-MM-DD/** : directory containing the reconstructed images generated on this date using the preprocessed data above.

  (a) **x_k.dcm** : DICOM images of x after k iterations.

  (b) **x_k.txt** : text images of x after k iterations.

  (c) **x_k.png** : PNG images of x after k iterations.

/data/ion/pCT_data/raw_data : directory containing the raw experimental data organized by scan date

(1) **/YY-MM-DD** : Folder containing all raw experimental data acquired from the scan beginning on "*YY-MM-DD*"

  (a) **<object>_XXXX[_AAA]_xxx.dat** : raw experimental data for the object named "*<object>*", from run # "*XXXX[_AAA]*", where "*XXXX*" is a 4 digit # with leading zeros and "*_AAA*" is an optional descriptor tag, and "*xxx*" is the gantry angle at which the data was acquired.

/data/ion/pCT_data/preprocessed_data : directory containing the preprocessed experimental data organized by scan and processed dates

(1) /YY-MM-DD/ : Folder containing all processed experimental data corresponding to the raw experimental data acquired on " *YY-MM-DD* "

    (a) /YY-MM-DD : Folder containing all processed experimental data generated on " *YY-MM-DD* " from the raw data

        (i) *<object>_XXXX[_AAA]_xxx.dat.root.reco.root.bin* : processed experimental data with tracker coordinates, recovery of missing hits when possible, and calibrated WEPL measurements for the object named " *<object>* ", from run # " *XXXX[_AAA]* ", where " *XXXX* " is a 4 digit # with leading zeros and " *_AAA* " is an optional descriptor tag, and " *xxx* " is the gantry angle at which the data was acquired.