



BAYLOR
UNIVERSITY

PCT CODE/DATA STORAGE AT BAYLOR

FOLDER/FILE NAMING AND ORGANIZATIONAL SCHEME
BLAKE EDWARD SCHULTZE

JULY 21, 2016

The logo for Baylor University Engineering & Computer Science. It features a green circle at the top left, a yellow swoosh that curves around the text, and a large, light gray ampersand in the background. The text 'Baylor University' is in a small, black, sans-serif font, and 'Engineering Computer Science' is in a larger, black, sans-serif font.

Baylor University
Engineering
Computer Science

TABLE OF CONTENTS

	Page
Directory Usage	2
Kodiak Directory Descriptions and Usage	2
Tardis Compute Node Directory Descriptions and Usage	5
Organizational Scheme	7
Code Organization	7
pCT Code Hierarchy	7
Data Organization	8
Organized Data Hierarchy	8
Raw Data Hierarchy	14
Preprocessed Data Hierarchy	14
Reconstruction Data Hierarchy	15
Unorganized User Data Hierarchy	20
GitHub Sources	21
GitHub Accounts/Repositories	21
Master File List	23

KEY:

Green : directories whose names do not change

Brown : directories w/ parameter dependent naming (e.g. object name, date, etc.)

Italic + **Royal Blue** : data/image file

Italic + **Dark Blue** : set of data/image files

1.1: KODIAK DIRECTORY DESCRIPTIONS AND USAGE

DESCRIPTION OF USER/SHARED DIRECTORIES ON KODIAK

KODIAK**/ion/...**

– This directory is dedicated to the storage of all files associated with proton and ion therapy research. There are private and shared subdirectories and to prevent inappropriate modifications/deletions of shared code/data, users only have write permissions to their private directories and they cannot directly add/modify shared data. To share data with the collaboration, users submit the data to their private “staging” directory using the appropriate naming/organization and an administrator will then validate the data and move it to the appropriate shared directory. Administrators typically will not be familiar with the naming/organizational scheme, but by organizing the data according to the full destination path, administrators can move the entire hierarchy of files/folders directly to **/ion**, thereby merging the contents of any existing directories automatically.

The data in this directory is located on a network storage device and can be accessed from all Kodiak and Tardis cluster nodes. The device is also backed up to tape drive periodically to prevent permanent loss of data in the event of drive failure.

/ion/home/<user name>/...

– This is a user’s private home directory where the files associated with their account are stored (e.g., **.bash_profile**, **.bash_history**, etc.) and is the default login directory. Each user only has access to their personal directory, but because it is on the network storage device, it can be accessed from each of the Kodiak/Tardis nodes. Now that the home directories have been moved to **/ion**, they no longer have a limited storage capacity, so users may run code and write the resulting output data/images to this directory. Note that as a subdirectory of **/ion**, the data in this directory will automatically be backed up to tape drive so it is recoverable in case of data corruption or drive failure.

`/data/<user name>/...`

- These private data directories can be used as an alternative to `/ion/home/<user name>` for storing input data for code/program execution and as the destination for the resulting output data generated. As subdirectories of `/data`, the contents of these directories are backed up periodically, so these can also be used for long term data storage.

`/ion/incoming/<user name>/...`

- These private directories are used to upload data to the Baylor server prior to moving it to the intended destination. When the uploaded data is intended to be shared with the collaboration, the directory should be used to rename and organize the data files according to the naming/organizational scheme before moving it to a user's private `/staging` directory, from which an administrator will validate and move the data to the appropriate shared directory.

`/ion/staging/<user name>/...`

- These directories are used to submit code/data for sharing with the collaboration. Since administrators are typically unfamiliar with the naming/organizational scheme for shared data, users must first rename/organize the data as needed to create the entire hierarchy of directories corresponding to the full destination path, including all subdirectories below `/ion`. Administrators need not know the destination path or understand the organization but can then simply move the entire hierarchy and the contents of existing directories such as `/ion` will automatically be merged and the new data/directories added. To simplify the creation of these hierarchies and ensure consistency by removing manual naming/organization, bash scripts/functions have been developed to organize data and move it to a user's `/staging` directory by passing the requisite information as execution parameters (e.g., phantom, run date/#/tag(s), etc.).

`/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_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.

/ion/pCT_code/...

– This directory is used to store permanent and semi-permanent pCT source code, from data acquisition to image reconstruction and analysis of reconstructed images. It contains clones of GitHub repositories as well as user’s personal versions of programs they want to make available to other users (otherwise users should keep their code in their private directories) organized by program type (Preprocessing/Reconstruction/etc.) with subdirectories for each user.

/ion/pCT_code/git/<GitHub account>/<GitHub repository>/...

– This directory contains clones of the available pCT GitHub accounts and repositories, with parent directories for each GitHub account and subdirectories for each of their repositories. Each program repository has a *master* branch, which typically corresponds to the current release version (though there may also be a branch like *release* used instead) and each of the program’s developers will typically have their own branch which they can use to develop and test new ideas/features. The group of developers of a program should decide amongst themselves what the process will be for approving merges with the *master/release* branch and when to release a new version of the program, which may include the results of several separate merges.

Users accessing the *master/release* branch of these clones should execute `git pull --rebase` prior to using the code to ensure it is updated to its latest version.

NOTE: This should not be done for other branches or the personal versions of code.

/ion/pCT_code/user_code/<user name>/...

– Contains subdirectories for each pCT user where they can store and modify clones of the pCT program repositories and their personal code.

1.2: TARDIS COMPUTE NODE DIRECTORY DESCRIPTIONS AND USAGE

DESCRIPTION OF USER/SHARED DIRECTORIES ON TARDIS COMPUTE NODES

TARDIS

`/local/...`

- This directory is the parent directory for all pCT code/data on each Tardis compute node, the equivalent of the `/ion` directory on Kodiak. This data is stored on the compute nodes' local solid-state drive and is not backed up, so users must ensure they copy data to Kodiak if they want to store it permanently

`/local/pCT_code/...`

- This directory is used to store pCT code on the Tardis compute nodes and corresponds to the `/ion/pCT_code` directory on Kodiak (with the same organizational scheme as well).

`/local/pCT_code/user_code/<user name>/...`

- The equivalent of Kodiak's `/ion/pCT_code/user_code/<user name>/...` subdirectories where pCT users can copy/modify/execute their personal clones of pCT program repositories and their personal code on Tardis compute nodes

`/local/pCT_code/git/<GitHub account>/<GitHub repository>/...`

- This directory contains clones of the available pCT GitHub accounts and repositories, with parent directories for each GitHub account and subdirectories for each of their repositories. Each program repository has a `master` branch, which typically corresponds to the current release version (though there may also be a branch like `release` used instead) and each of the program's developers will typically have their own branch which they can use to develop and test new ideas/features. The group of developers of a program should decide amongst themselves what the process will be for approving merges with the `master/release` branch and when to release a new version of the program, which may include the results of several separate merges.

Users accessing the `master/release` branch of these clones should execute `git pull --rebase` prior to using the code to ensure it is updated to its latest version. NOTE: This should not be done for other branches or the personal versions of code.

`/local/pCT_data/...`

- This directory is where pCT data is to be copied from Kodiak and corresponds to the `/ion/pCT_data` directory on Kodiak.

`/local/pCT_data/user_data/<user name>/...`

- Subdirectories for each pCT user on each Tardis compute node where they can transfer data they want to reconstruct when the data is not organized according to the pCT data naming/organizational scheme. If the output data they generate will also be unorganized, it should also be written to this directory.

NOTE: The `pCT-collaboration/pCT_Tools` repository contains a script which loads a number of bash functions useful to pCT users, including a function that can be used to organize and then copy unorganized data to the organized data directory on the Tardis compute nodes

2.1: CODE ORGANIZATION

2.1 pCT CODE HIERARCHY

pCT_CODE

`/ion/local/pCT_code/...`

– directory containing pCT code on Kodiak and Tardis compute nodes, both of which have the same organizational scheme, the only difference being the top level parent directory on Kodiak is `/ion` and on the Tardis compute nodes is `/local`.

- (1) `/ion` or `/local` : parent directory for all pCT code/data on Kodiak and the Tardis compute nodes
 - (a) `/pCT_code` : directory containing the code for all pCT programs linked to their GitHub repositories as well as subdirectories for each pCT user where they can clone and modify these repositories and store/execute their own code
 - (i) `/git` : directory containing clones of the standard/common pCT programs, providing easy and immediate access to the newest version of each of these programs
 - (1) `/<GitHub account>` : directories for each of the GitHub accounts containing one or more pCT programs
 - (a) `/<GitHub repository>` : subdirectories for each pCT code repository in the associated GitHub account
 - (ii) `/user_code` : directory containing subdirectories for each user where they can store their personal code
 - (1) `/<username>` : subdirectories for each pCT user where they can store their personal code

2.2: DATA ORGANIZATION

2.2 ORGANIZED DATA HIERARCHY

ORGANIZED DATA

/ion/pCT_data/organized_data

– directory containing all of the pCT data (raw, preprocessed/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 this phantom/object.
 - (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 experimental scans of the object.
 - (i) **/YY-MM-DD** : directory containing data and reconstructed images corresponding to the experimental scan of the object performed on this date.
 - (1) **/XXXX[_AAA]** : directory containing data and reconstructed images corresponding to the 4-digit run # “XXXX”, potentially including “*subcategory tag(s)*” of the form “_AAA” such as “_Cont” specifying a continuous scan and/or phantom positioning tags such as inferior (“_Inf”) or superior (“_Sup”), top (“_Top”) or bottom (“_Bot”), etc.
 - (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) **readme.txt** : contains input raw data info, phantom name, and run date.
 - (2) **TVcorr.txt** : contains TV corrected WEPL calibration curve coefficients.
 - (3) **WcalibTemp.txt** : temporary file containing WEPL calibration curve coefficients.
 - (4) **Wcalib.txt** : contains final WEPL calibration curve coefficients.

- (5) *projection.xxx.bin* : preprocessed data files containing tracker coordinates and WEPL data for gantry angle “xxx” = {“001”, “002”, “003”, ...} used as input to image reconstruction.
- (6) */Reconstruction* : directory containing preprocessed data and reconstructed images generated using the “*projection.xxx.bin*” data along with reference images relevant to the object.
 - (a) *settings.cfg* : specifies scan properties such as gantry angle interval, t/v detector size, reconstruction volume dimensions, etc and default settings and parameters to use in reconstructing this data set.
 - (b) */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_log.txt* : execution times for various portions of preprocessing and/or reconstruction and total program execution time.
 - (iii) *TV_measurements.txt* : contains total variation (TV) measurements before/after each iteration
 - (iv) *hull.txt* : text image of object hull in 1s/0s with the xy plane for each slice stacked on each other.
 - (v) *hull.png* : “*hull.txt*” converted to PNG image.
 - (vi) *FBP.txt* : text image of filtered back projection (FBP) image with the xy plane for each slice stacked on each other.
 - (vii) *FBP.png* : “*FBP.txt*” converted to PNG image.
 - (viii) *FBP_med_filtered.txt* : text image result of applying median filter to the filtered back projection (FBP) image with the xy plane for each slice stacked on each other.
 - (ix) *FBP_med_filtered.png* : “*FBP_avg_filtered.txt*” converted to PNG image.
 - (x) *FBP_avg_filtered.txt* : text file specifying the result of applying an average filter to the filtered back projection image with the xy plane for each slice stacked on each other.
 - (xi) *FBP_avg_filtered.png* : “*FBP_avg_filtered.txt*” converted to PNG image.
 - (xii) *x_0.txt* : text image of initial iterate with the xy plane for each slice stacked on each other.
 - (xiii) *x_0.png* : “*x_0.txt*” converted to PNG image.
 - (xiv) */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.txt* : text images of x after k iterations.
 - (b) *x.k.png* : PNG images of x after k iterations.
 - (c) **/Simulated** : directory containing data and images generated from simulated scans of the object.
 - (i) **/G.YY-MM-DD** : directory containing data and reconstructed images from the GEANT4 simulated scan of the object generated on this date..
 - (1) **/XXXX[AAA]** : directory containing data and reconstructed images corresponding to this run # of the scan date, where the run # is of the form “XXXX” with optional subcategory tag(s) “_AAA” specifying additional pertinent information about the scan, such as inferior “_Inf” or superior “_Sup” positioning (e.g. for 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) *readme.txt* : contains input raw data info, phantom name, and run date.
 - (2) *TVcorr.txt* : contains TV corrected WEPL calibration curve coefficients.
 - (3) *WcalibTemp.txt* : temporary file containing WEPL calibration curve coefficients.
 - (4) *Wcalib.txt* : contains final WEPL calibration curve coefficients.
 - (5) *projection.xxx.bin* : preprocessed data files containing tracker coordinates and WEPL data for gantry angle “xxx” = {“001”, “002”, “003”, ...} used as input to image reconstruction.
 - (6) **/Reconstruction** : directory containing preprocessed data and reconstructed images generated using the “*projection.xxx.bin*” data along with reference images relevant to the object.
 - (a) *settings.cfg* : specifies scan properties such as gantry angle interval, t/v detector size, reconstruction volume dimensions, etc and default settings and parameters to use in reconstructing this data set.
 - (b) **/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 pa-

- rameters/options applied at execution, if any.
- (ii) *execution_log.txt* : execution times for various portions of preprocessing and/or reconstruction and total program execution time.
 - (iii) *TV_measurements.txt* : contains total variation (TV) measurements before/after each iteration
 - (iv) *hull.txt* : text file specifying hull in 1s/0s with the xy plane for each slice stacked on each other.
 - (v) *hull.png* : “*hull.txt*” converted to PNG image.
 - (vi) *FBP.txt* : text file specifying filtered back projection image with the xy plane for each slice stacked on each other.
 - (vii) *FBP.png* : “*FBP.txt*” converted to PNG image.
 - (viii) *FBP_med_filtered.txt* : text file specifying the result of applying a median filter to the filtered back projection image with the xy plane for each slice stacked on each other.
 - (ix) *FBP_med_filtered.png* : “*FBP_avg_filtered.txt*” converted to PNG image.
 - (x) *FBP_avg_filtered.txt* : text file specifying the result of applying an average filter to the filtered back projection image with the xy plane for each slice stacked on each other.
 - (xi) *FBP_avg_filtered.png* : “*FBP_avg_filtered.txt*” converted to PNG image.
 - (xii) *x_0.txt* : text file specifying voxel values of initial iterate with the xy plane for each slice stacked on each other.
 - (xiii) *x_0.png* : text file specifying voxel values of initial iterate with the xy plane for each slice stacked on each other.
 - (xiv) */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.txt* : text images of x after k iterations.
 - (b) *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 this run # of the scan date, where the run # is of the form “XXXX” with optional subcategory tag(s) “_AAA” specifying additional pertinent information about the scan, such as inferior “_Inf” or superior “_Sup” positioning (e.g. for 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) **readme.txt** : contains input raw data info, phantom name, and run date.
 - (2) **TVcorr.txt** : contains TV corrected WEPL calibration curve coefficients.
 - (3) **WcalibTemp.txt** : temporary file containing WEPL calibration curve coefficients.
 - (4) **Wcalib.txt** : contains final WEPL calibration curve coefficients.
 - (5) **projection.xxx.bin** : preprocessed data files containing tracker coordinates and WEPL data for gantry angle “xxx” = {“001”, “002”, “003”, ...} used as input to image reconstruction.
 - (6) **/Reconstruction** : directory containing preprocessed data and reconstructed images generated using the “**projection.xxx.bin**” data along with reference images relevant to the object.
 - (a) **settings.cfg** : specifies scan properties such as gantry angle interval, t/v detector size, reconstruction volume dimensions, etc and default settings and parameters to use in reconstructing this data set.
 - (b) **/YY-MM-DD** : directory containing the pre-reconstruction processed 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_log.txt** : execution times for various portions of preprocessing and/or reconstruction and total program execution time.
 - (iii) **TV_measurements.txt** : text file containing total variation (TV) measurements before/after each iteration
 - (iv) **hull.txt** : text file specifying hull in 1s/0s with the xy plane for each slice stacked on each other.
 - (v) **hull.png** : “**hull.txt**” converted to PNG image.
 - (vi) **FBP.txt** : text file specifying filtered back projection image with the xy plane for each slice stacked on each other.
 - (vii) **FBP.png** : “**FBP.txt**” converted to PNG image.
 - (viii) **FBP_med_filtered.txt** : text file specifying the result of applying a me-

- dian filter to the filtered back projection image with the xy plane for each slice stacked on each other.
- (ix) *FBP_med_filtered.png* : “*FBP_avg_filtered.txt*” converted to PNG image.
 - (x) *FBP_avg_filtered.txt* : text file specifying the result of applying an average filter to the filtered back projection image with the xy plane for each slice stacked on each other.
 - (xi) *FBP_avg_filtered.png* : “*FBP_avg_filtered.txt*” converted to PNG image.
 - (xii) *x_0.txt* : text file specifying voxel values of initial iterate with the xy plane for each slice stacked on each other.
 - (xiii) *x_0.png* : text file specifying voxel values of initial iterate with the xy plane for each slice stacked on each other.
 - (xiv) */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.txt* : text images of x after k iterations.
 - (b) *x_k.png* : PNG images of x after k iterations.

2.3 RAW DATA HIERARCHY

RAW_DATA

/ion/pCT_data/raw_data

– directory where new raw experimental data files for all objects scanned on a particular scan date prior to creation of soft links renamed to “*projection.xxx.bin*” and organized according to the naming/organizational scheme.

- (1) /YY-MM-DD : Folder containing all raw experimental data acquired from the scan beginning on “YY-MM-DD”
 - (a) <Phantom>_XXXX[_AAA].xxx.dat : raw experimental data for the object named “<Phantom>”, 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.

2.4 PREPROCESSED DATA HIERARCHY

PREPROCESSED_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) <Phantom>_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 “<Phantom>”, 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.

2.5 RECONSTRUCTION DATA HIERARCHY

RECONSTRUCTION DATA

/ion/pCT_data/reconstruction_data

– directory containing the reconstructed images/data. Although the raw and preprocessed/projection data is not stored in this directory as well, it is organized into the same hierarchy of subdirectories as the organized data to make the data dependencies clear.

- (1) **/Phantom** : directory containing all of the experimental/simulated data and reconstructed images associated with this phantom/object.
 - (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 the experimental scan of the object performed on this date.
 - (1) **/XXXX[_AAA]** : directory containing data and reconstructed images corresponding to this run # of the scan date, where the run # is of the form “XXXX” with optional subcategory tag(s) “_AAA” specifying additional pertinent information about the scan, such as inferior “_Inf” or superior “_Sup” positioning (e.g. for 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) **projection.xxx.bin** : preprocessed data files containing tracker coordinates and WEPL data for gantry angle “xxx” = {“001”, “002”, “003”, ...} 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) *settings.cfg* : specifies scan properties such as gantry angle interval, t/v detector size, reconstruction volume dimensions, etc and default settings and parameters to use in reconstructing this data set.
- (b) */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.log.txt* : execution times for various portions of preprocessing and/or reconstruction and total program execution time.
 - (iii) *TV-measurements.txt* : text file containing total variation (TV) measurements before/after each iteration
 - (iv) *hull.txt* : text file specifying hull in 1s/0s with the xy plane for each slice stacked on each other.
 - (v) *hull.png* : “*hull.txt*” converted to PNG image.
 - (vi) *FBP.txt* : text file specifying filtered back projection image with the xy plane for each slice stacked on each other.
 - (vii) *FBP.png* : “*FBP.txt*” converted to PNG image.
 - (viii) *FBP_med_filtered.txt* : text file specifying the result of applying a median filter to the filtered back projection image with the xy plane for each slice stacked on each other.
 - (ix) *FBP_med_filtered.png* : “*FBP_avg_filtered.txt*” converted to PNG image.
 - (x) *FBP_avg_filtered.txt* : text file specifying the result of applying an average filter to the filtered back projection image with the xy plane for each slice stacked on each other.
 - (xi) *FBP_avg_filtered.png* : “*FBP_avg_filtered.txt*” converted to PNG image.
 - (xii) *x_0.txt* : text file specifying voxel values of initial iterate with the xy plane for each slice stacked on each other.
 - (xiii) *x_0.png* : text file specifying voxel values of initial iterate with the xy plane for each slice stacked on each other.
 - (xiv) */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.txt* : text images of x after k iterations.
 - (b) *x_k.png* : PNG images of x after k iterations.
- (c) */Simulated* : directory containing data and images generated from simulated scans of the object.
 - (i) */G_YY-MM-DD* : directory containing data and reconstructed images from the GEANT4 simulated

scan of the object generated on this date..

- (1) **/XXXX[_AAA]** : directory containing data and reconstructed images corresponding to this run # of the scan date, where the run # is of the form “XXXX” with optional subcategory tag(s) “_AAA” specifying additional pertinent information about the scan, such as inferior “_Inf” or superior “_Sup” positioning (e.g. for 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** : preprocessed data files containing tracker coordinates and WEPL data for gantry angle “xxx” = {“001”, “002”, “003”, ...} 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) **settings.cfg** : specifies scan properties such as gantry angle interval, t/v detector size, reconstruction volume dimensions, etc and default settings and parameters to use in reconstructing this data set.
 - (b) **/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.log.txt** : execution times for various portions of preprocessing and/or reconstruction and total program execution time.
 - (iii) **TV-measurements.txt** : text file containing total variation (TV) measurements before/after each iteration
 - (iv) **hull.txt** : text file specifying hull in 1s/0s with the xy plane for each slice stacked on each other.
 - (v) **hull.png** : “**hull.txt**” converted to PNG image.
 - (vi) **FBP.txt** : text file specifying filtered back projection image with the xy plane for each slice stacked on each other.

- (vii) *FBP.png* : “*FBP.txt*” converted to PNG image.
- (viii) *FBP_med_filtered.txt* : text file specifying the result of applying a median filter to the filtered back projection image with the xy plane for each slice stacked on each other.
- (ix) *FBP_med_filtered.png* : “*FBP_avg_filtered.txt*” converted to PNG image.
- (x) *FBP_avg_filtered.txt* : text file specifying the result of applying an average filter to the filtered back projection image with the xy plane for each slice stacked on each other.
- (xi) *FBP_avg_filtered.png* : “*FBP_avg_filtered.txt*” converted to PNG image.
- (xii) *x_0.txt* : text file specifying voxel values of initial iterate with the xy plane for each slice stacked on each other.
- (xiii) *x_0.png* : text file specifying voxel values of initial iterate with the xy plane for each slice stacked on each other.
- (xiv) */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.txt* : text images of x after k iterations.
 - (b) *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 this run # of the scan date, where the run # is of the form “XXXX” with optional subcategory tag(s) “_AAA” specifying additional pertinent information about the scan, such as inferior “_Inf” or superior “_Sup” positioning (e.g. for 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* : preprocessed data files containing tracker coordinates and

WEPL data for gantry angle “ xxx ” = {“001”, “002”, “003”, ...} 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) *settings.cfg* : specifies scan properties such as gantry angle interval, t/v detector size, reconstruction volume dimensions, etc and default settings and parameters to use in reconstructing this data set.
 - (b) /**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_log.txt* : execution times for various portions of preprocessing and/or reconstruction and total program execution time.
 - (iii) *TV_measurements.txt* : text file containing total variation (TV) measurements before/after each iteration
 - (iv) *hull.txt* : text file specifying hull in 1s/0s with the xy plane for each slice stacked on each other.
 - (v) *hull.png* : “*hull.txt*” converted to PNG image.
 - (vi) *FBP.txt* : text file specifying filtered back projection image with the xy plane for each slice stacked on each other.
 - (vii) *FBP.png* : “*FBP.txt*” converted to PNG image.
 - (viii) *FBP_med_filtered.txt* : text file specifying the result of applying a median filter to the filtered back projection image with the xy plane for each slice stacked on each other.
 - (ix) *FBP_med_filtered.png* : “*FBP_avg_filtered.txt*” converted to PNG image.
 - (x) *FBP_avg_filtered.txt* : text file specifying the result of applying an average filter to the filtered back projection image with the xy plane for each slice stacked on each other.
 - (xi) *FBP_avg_filtered.png* : “*FBP_avg_filtered.txt*” converted to PNG image.
 - (xii) *x_0.txt* : text file specifying voxel values of initial iterate with the xy plane for each slice stacked on each other.
 - (xiii) *x_0.png* : text file specifying voxel values of initial iterate with the xy plane for each slice stacked on each other.
 - (xiv) /**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.txt* : text images of x after k iterations.
 - (b) *x.k.png* : PNG images of x after k iterations.

2.6 UNORGANIZED USER DATA HIERARCHY

USER_DATA

/ion/pCT_data/user_data

– directory containing unorganized input and output reconstruction data, allowing users to use and keep their unorganized data separate from other data and maintain it in their preferred organizational scheme without it interfering with the properly organized data.

- (1) **/user_data** : directory unique to Tardis compute nodes containing subdirectories for each user where they can transfer unorganized data they want to reconstruct and write the corresponding output reconstruction data/images
 - (a) **/<username>** : subdirectories for each pCT user for the unorganized input and output reconstruction data

3.1: GITHUB ACCOUNTS/REPOSITORIES

GITHUB

<https://github.com/<GitHub account>/<GitHub repository>>

– below is an outline of the GitHub accounts and repositories containing documentation, tools, and code relevant to the storage of pCT data/code, data acquisition, processing, and analysis, and the various programs and tools used to perform pCT related tasks

- (1) **/pCT-collaboration** : parent directory for all pCT code/data on Kodiak and the Tardis compute nodes
 - (a) **/pCT_Tools** : contains bash functions/scripts and other tools useful for navigating data/code and configuring/running programs on Kodiak and Tardis compute nodes along with documentation describing them and their purpose/usage. There is also a default `.bash_profile` for pCT users which sources `pct_user_script.sh` to configure user sessions depending on the current Kodiak/Tardis node and `load_pct_functions.sh` to automatically load the aforementioned bash functions useful during a user terminal session.
 - (b) **/pCT-docs** : contains documentation on the pCT data/code naming and organizational scheme, collaborator's project involvement and contact info, and phantom naming (including relevant subcategory tags) and properties/manuals. The contents of this repository are primarily being added from the original **/BlakeSchultze/pCT_Documentation** documentation repository, but the naming and contents of these files are being verified/improved prior to adding them to this repository and some of the original files may not be added. This repository is currently incomplete.
 - (c) **/pypct** : Python helpers for proton CT
 - (d) **/pct-acquire** : parent directory for all pCT code/data on Kodiak and the Tardis compute nodes
 - (e) **/pct-sim** : Geant4 simulations for proton CT
 - (f) **/Preprocessing** : program for preprocessing raw data to generate tracker coordinates and perform WEPL calibration.
 - (g) **/pct-recon-copy** : original pCT reconstruction program developed by Penfold/Hurley, now including separate branches for execution on Tardis compute nodes using modified Makefile and addition of execution parameters for input/output directory allowing these to be specified at runtime without requiring recompilation, a capability required for development of a batch script to submit reconstruction job(s) to the GPU execution queue.
 - (h) **/Reconstruction_BU** : contains only the current and previous release versions of Baylor's reconstruction program as developed in **/BaylorICTHUS/pCT_Reconstruction** (no code development is performed here).

- (2) **/BaylorICTHUS** : Baylor’s pCT programs, tools, and documentation.
 - (a) **/pCT_Reconstruction** : used in developing the release version of Baylor’s pCT reconstruction program and containing branches for each of Baylor’s pCT developers (Blake, Paniz, Sarah, ...) for independent development relevant to their work. Developments made in a developer branch and proposed for integration in the next release version go through a review and testing process to verify the code and its impact on the full program. Developments passing this verification process are then merged into the **release_development** branch. When critical developments are merged into the **release_development** branch, this branch is then merged into the **release** branch and the resulting code is then pushed to the **/pCT-collaboration/Reconstruction.BU** repository, as this is the source for pCT users to acquire the current and previous release versions of the program.
- (3) **/BlakeSchultze** : parent directory for all pCT code/data on Kodiak and the Tardis compute nodes
 - (a) **/LaTeX-Packages** : provides the package “**my-latex.sty**” which is included in TeX documents to provide access to the definitions of new commands/macros/environments, load the external/3rd-party package dependencies, and configure the typesetting of LaTeX documents as well as providing the collection of LaTeX style (.sty) and other files included in this repository upon which these definitions/configurations are dependent.
 - (b) **/pCT_Documentation** : contains an expanded set of pCT documentation files with additional resources not included in the **/pCT-collaboration/pCT-docs** repository, such as pCT publications and theses/dissertations.
 - (c) **/pCT_Reconstruction** : the original repository in which Baylor’s pCT reconstruction program was developed, which also contains the experimental development of an alternative program configuration with several automated routines, and is currently being merged into the release version of Baylor’s reconstruction program as provided in **/BaylorICTHUS/pCT_Reconstruction**.
 - (d) **/WED_Analysis** : provides tool for determining the water-equivalent depth (WED) for a set of beam-aim point (BAP) coordinates based on reconstructed image RSP values, using the voxel walk algorithm developed as part of the pCT reconstruction program. This algorithm steps from voxel edge to voxel edge along a trajectory to determine exact voxel intersection coordinates and prevent the missing of small voxel intersections which can occur when taking constant length steps along a path as was done in the original reconstruction program.

MASTER FILE LIST

Contents:

BELOW IS A MASTER LIST OF FILES AND THEIR DESCRIPTIONS IN ALPHABETICAL ORDER.

1. *<Phantom>_XXXX[_AAA]_xxx.dat* : raw experimental data for the object named “<Phantom>”, 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.
2. *<Phantom>_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 “<Phantom>”, 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.
3. *bin_counts.txt* : linearized bin # for each proton history, where linearized bin # = $t_bin + angle_bin * T_BINS + v_bin * T_BINS * ANGULAR_BINS$.
4. *calib.txt* : text file containing calibration curve coefficients for WEPL calibration.
5. *coefficient.bin* : file containing the tabulated scattering coefficient values for Σ_1/Σ_2 for $u_2 - u_1/u_1$ values
6. *cos_table.bin* : file containing the tabulated values of cosine function
7. *execution_log.csv* : The same information listed in the *execution_log.txt* files for each execution are also stored in this global comma separated value (.csv) file with the execution information organized into separate columns and each row corresponding to a different reconstruction, thereby maintaining a single file where the execution information for all previous reconstructions can be found in the same place. Upon execution of a new reconstruction, a new row entry is added for storage of the corresponding execution information.
8. *execution_log.txt* : Listing of the preprocessing/reconstruction options/parameters used for a particular reconstruction along with the execution times for various individual tasks and portions of preprocessing and/or reconstruction and total program execution time organized as a series of “key=value” entries in a text file in the directory with the corresponding results.
9. *FBP.txt* : text file specifying filtered back projection image with the xy plane of each slice stacked on each other.

10. *FBP.png* : “*FBP.txt*” converted to PNG image.
11. *FBP_hull.txt* : text file specifying object hull generated from the filtered back projection image by setting all voxels with RSP less than a user definable threshold (typically 0.6) with the xy plane of each slice stacked on each other.
12. *FBP_hull.png* : “*FBP.txt.png*” converted to PNG image
13. *FBP_med_filtered.txt* : text file specifying the result of applying a median filter to the filtered back projection image with the xy plane for each slice stacked on each other.
14. *FBP_med_filtered.png* : “*FBP.txt*” converted to PNG image.
15. *FBP_avg_filtered.txt* : text file specifying the result of applying an average filter to the filtered back projection image with the xy plane for each slice stacked on each other.
16. *FBP_avg_filtered.txt* : “*FBP_avg_filtered.txt*” converted to PNG image.
17. *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.
18. *hull.txt* : text file specifying hull in 1s/0s with the xy plane of each slice stacked on each other.
19. *hull.png* : “*hull.txt*” converted to PNG image.
20. *hull_avg_filtered.txt* : text file specifying hull after applying an average filter and applying a threshold to the result so that all voxels with value less than some threshold (typically 0.01-0.1) are set to 0, thereby restoring image to 1s/0s with the xy plane of each slice stacked on each other.
21. *mean_rel_ut_angle.txt* : mean relative ut angle ($\angle_{ut_{out}} - \angle_{ut_{in}}$) by linearized bin #.
22. *mean_rel_uv_angle.txt* : mean relative uv angle ($\angle_{uv_{out}} - \angle_{uv_{in}}$) by linearized bin #.
23. *mean_WEPL.txt* : mean WEPL value by linearized bin #.
24. *MLP.bin* : binary file with MLP path data for each history entering hull.
25. *MSC_counts.txt* : text file specifying Modified Space/Silhouette Carving generated image of integer values indicating the # of times each voxel was identified as lying outside the object with the xy plane of each slice stacked on each other.
26. *MSC_hull.txt* : text file specifying Modified Space/Silhouette Carving generated hull in 1s/0s with the xy plane of each slice stacked on each other.
27. *poly_1_2.bin* : file containing the tabulated MLP polynomial values with coefficients {1, 2, 3, 4, 5, 6}
28. *poly_2_3.bin* : file containing the tabulated MLP polynomial values with coefficients {2, 3, 4, 5, 6, 7}

29. *poly_3_4.bin* : file containing the tabulated MLP polynomial values with coefficients $\{3, 4, 5, 6, 7, 8\}$
30. *poly_2_6.bin* : file containing the tabulated MLP polynomial values with coefficients $\{2, 6, 12, 20, 30, 42\}$
31. *poly_3_12.bin* : file containing the tabulated MLP polynomial values with coefficients $\{3, 12, 30, 60, 105, 168\}$
32. *projection_xxx.bin* : binary files containing tracker coordinates and WEPL data associated with gantry position “xxx” = $\{“001”, “002”, “003”, \dots\}$ converted from raw data using WEPL calibration routine and used as input to image reconstruction.
33. *raw_xxx.bin* : binary files containing trigger/tracker/energy detector data from event builder associated with gantry position “xxx” = $\{“001”, “002”, “003”, \dots\}$.
34. *settings_log.cfg* : copy of “*settings.cfg*” with any changes made to parameters/options applied at execution, if any.
35. *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.
36. *sin_table.bin* : file containing the tabulated values of sine function
37. *stddev_rel_ut_angle.txt* : standard deviation of the relative ut angle ($\angle ut_{out} - \angle ut_{in}$) by linearized bin #.
38. *stddev_rel_uv_angle.txt* : standard deviation of the relative uv angle ($\angle uv_{out} - \angle uv_{in}$) by linearized bin #.
39. *stddev_WEPL.txt* : standard deviation of the WEPL value by linearized bin #.
40. *TVcalib.txt* : text file containing calibration curve coefficients for TV corrected WEPL calibration.
41. *TV_measurements.txt* : text file containing total variation (TV) measurements before/after each iteration.
42. *x_0.txt* : text file specifying voxel values of initial iterate with the xy plane of each slice stacked on each other.
43. *x_0.png* : “*x_0.txt*” converted to PNG image.
44. *x.k.txt* : text images of x after k iterations.
45. *x.k.png* : PNG images of x after k iterations.
46. *WEPL.bin* : binary file specifying WEPL value for each history entering hull.