

Version
0.1

MRC, HARWELL

Systems Imaging Group

HARP UserManual

Table of Contents

Introduction	1
Getting Started	2
Windows	2
Linux	2
How to use HARP	3
Overview	3
Imaging Modality Selection	4
Input and Output Select	5
Recon File Info	6
Identification	6
μ CT Specific Fields	7
Options	8
Status section	10
Add to Processing List	10
Output Summary	11
Error and Warning Messages	13
Appendix	15

Introduction

This manual will give a step by step guide on how to use HARP (Harwell Automated Recon Processor). The tool has been developed by the Systems Imaging Group at MRC Harwell as an aid in the processing of μ CT and OPT reconstructed images.

The intended end user is an employee working on the IMPC embryo phenotyping pipeline.

HARP is currently in a testing Beta stage and all functionality has not yet been implemented. Version 0.1 of the software is capable of cropping, scaling and compressing μ CT data.

For any queries please contact any of the following:

Henrik Westerberg: h.westerberg@har.mrc.ac.uk

Neil Horner: n.horner@har.mrc.ac.uk

Tom Lawson: t.lawson@har.mrc.ac.uk

Getting Started

HARP has been developed to work in either Linux or Windows environments. Below describes the steps to get HARP up and running.

Windows

HARP can either be run from an executable (advised) or run as a python script.

To run HARP from a Windows executable simply copy the latest version of HARP into the desired location. Then navigate to the Main.exe file and double click on the file. Alternatively a shortcut can be made by right clicking on the Main.exe file, selecting “create shortcut”. This shortcut can be copied to a convenient location.

Box 1: HARP executable location

```
C:\soft\HARP\dist\Main.exe
```

To run HARP directory from Python script a number of python modules will need to be installed. See Appendix for full list. Once the modules are installed, copy the HARP scripts to the desired location and run the Main.py in python.

Linux

HARP can currently be run as a python script in Linux. The dependent modules need to be installed. Once the modules are installed, copy the HARP scripts into the desired location and run the Main.py in python.

Box 2: Run HARP as Python Script

```
python Main.py
```

Section 3

How to use HARP

This section will take you through each step of how to process your data with HARP. The screenshots are taken from the Windows version of the software but apart from small variations in appearance all information should be relevant.

Overview

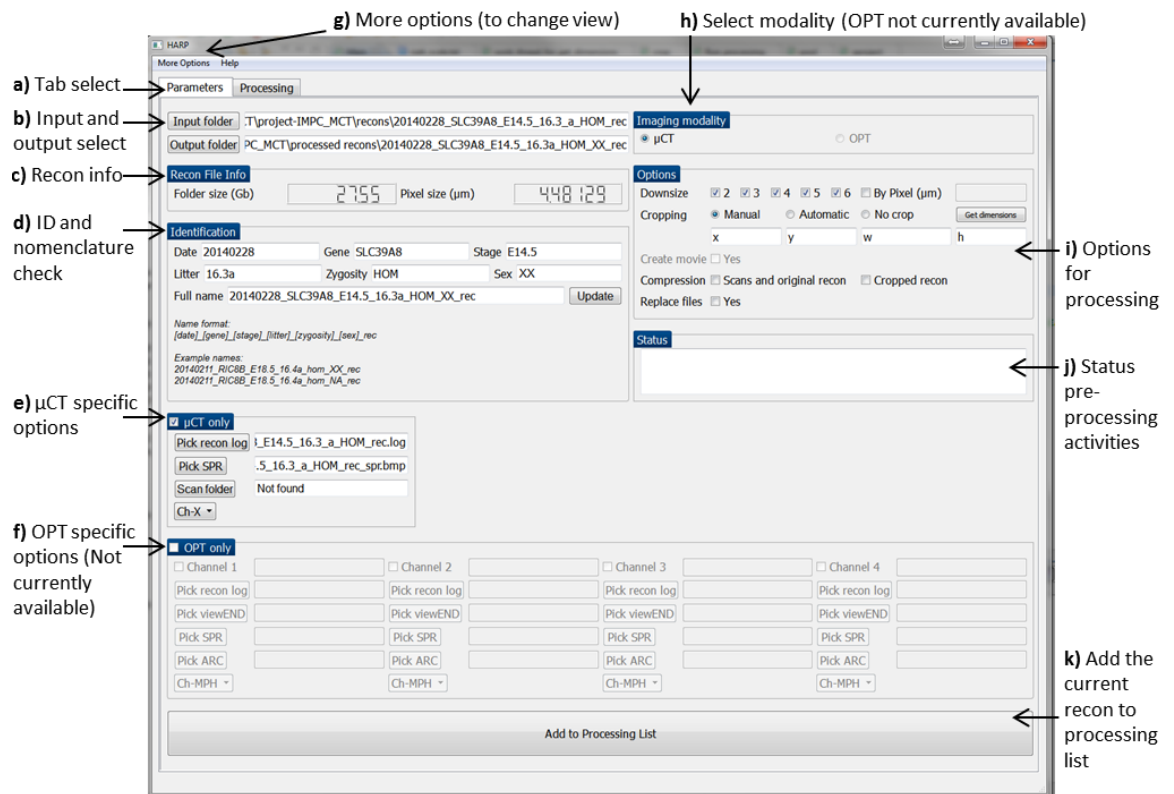


FIGURE 1: Starting view of HARP. The current tab is selected on "Parameters". Currently the any functionality not currently available is grey out and the user at this point cannot use. NOTE: See pdf version of user guide for higher resolution image.

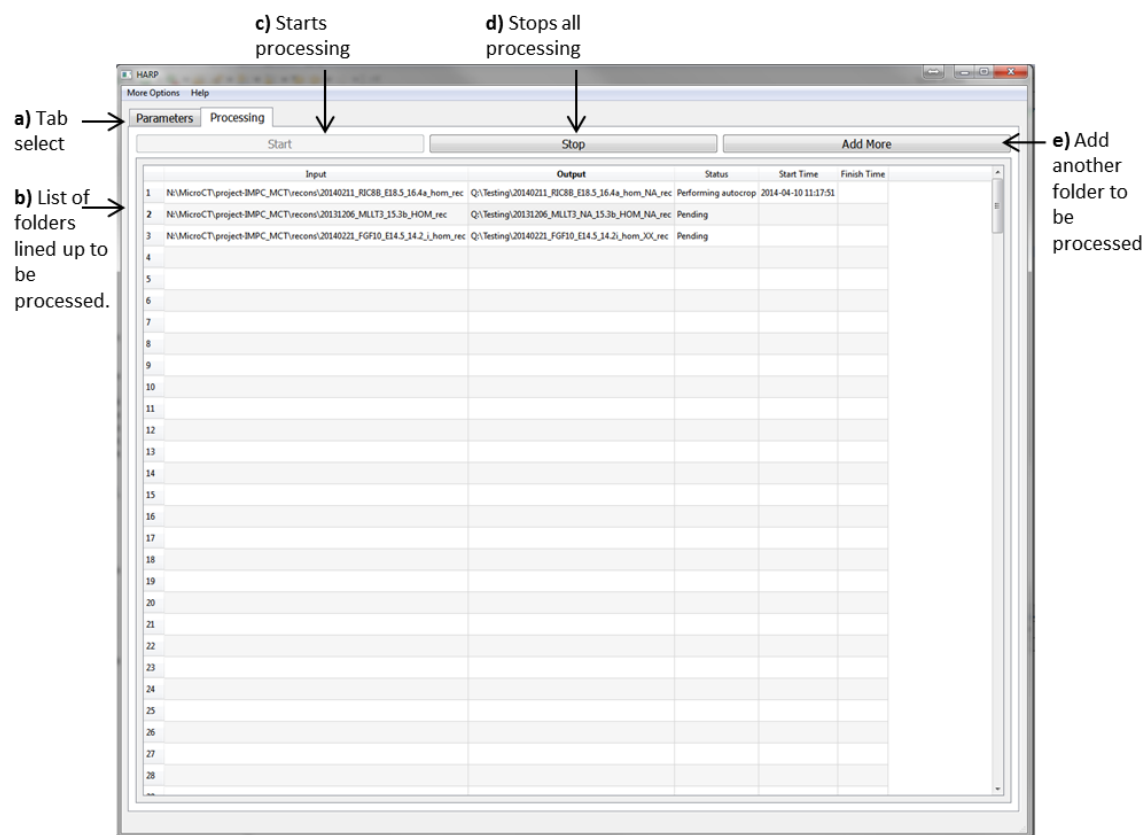


FIGURE 2: Processing view of HARP, shown when the “Processing” tab is selected and whenever as user successfully adds a recon folder to the processing list. NOTE: See pdf version of manual for higher resolution image.

Imaging Modality Selection

HARP is currently only available for μ CT image data. As such the imaging modality is locked on μ CT.

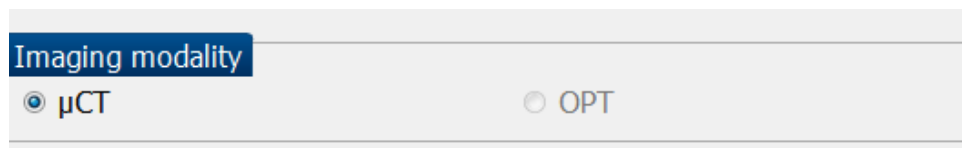


FIGURE 3: Imaging modality selection.

Input and Output Select

The first step in any use of the HARP will be to select the files to be processed. This is achieved by selecting the input folder button. The user should select the input directory which contains all reconstructed images which are required for processing.

After a folder have been selected HARP attempts to auto-populate a number of fields. This includes suggesting an output folder, calculating the recon folder size, identifying the pixel size, checking the identification nomenclature, locating the recon log file, locating the SPR file and locating the relevant scan folder.

If the folder being processed is from the IMPC pipeline, the output folder is automatically assigned to the “processed recons” directory in the project-IMPC_MCT directory. Alternatively the output folder can be assigned manually by clicking on the “output folder” button

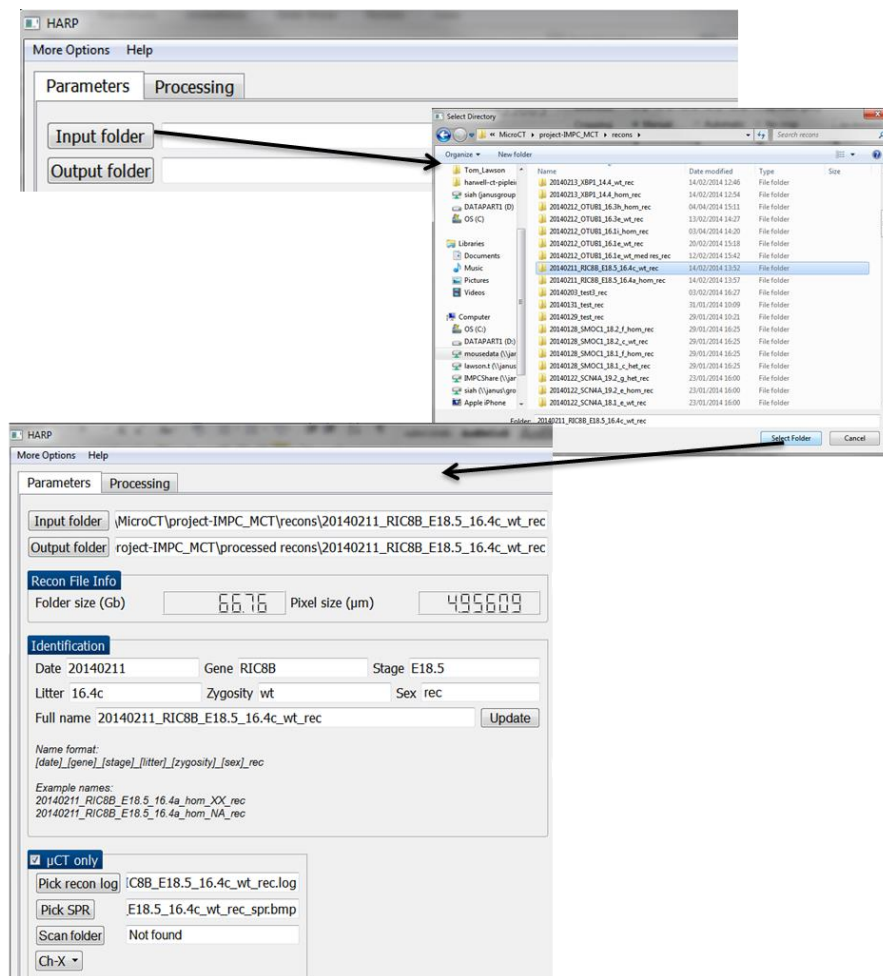


FIGURE 4: Steps to select input folder and the results of the auto-population of fields.
NOTE: See pdf version of user guide for higher resolution image.

Recon File Info

The recon file info is updated when an input folder is selected.

This section lets the user know what the folder size is of the original recon directory. The folder size calculation is based on only the recon files and does not include any subdirectories.

The pixel size is determined from the associated recon log file.

Recon File Info

Folder size (Gb)66.76Pixel size (µm)495609

FIGURE 5: Recon file information.

Identification

The identification section allows the user to see if the nomenclature of the reconstruction is in a standard format. The format includes the information date, gene, stage, litter, zygoty and sex separated by an underscore. If one of the identifiers is not known the letters NA should take its place. See below.

Box 3: Recon Nomenclature

Name format:

[date]_[gene]_[stage]_[litter]_[zygoty]_[sex]_rec

Example names:

20140211_RIC8B_E18.5_16.4a_hom_XX_rec

20140211_RIC8B_E18.5_16.4a_hom_NA_rec

If the identifiers do not fit into the desired format they should be updated in the “Full name” section and then the button “Update” should be clicked. See Figure 5 for example.

a

Identification			
Date	20140211	Gene	RIC8B
Stage	E18.5	Litter	16.4c
Zygoty	wt	Sex	rec
Full name 20140211_RIC8B_E18.5_16.4c_wt_rec			
Update			

Name format:
[date]_[gene]_[stage]_[litter]_[zygoty]_[sex]_rec

Example names:
20140211_RIC8B_E18.5_16.4a_hom_XX_rec
20140211_RIC8B_E18.5_16.4a_hom_NA_rec

Explanation of desired nomenclature

Where to edit name

Press to update fields

b

Identification			
Date	20140211	Gene	RIC8B
Stage	E18.5	Litter	16.4c
Zygoty	wt	Sex	XX
Full name 20140211_RIC8B_E18.5_16.4c_wt_XX_rec			
Update			

Name format:
[date]_[gene]_[stage]_[litter]_[zygoty]_[sex]_rec

Example names:
20140211_RIC8B_E18.5_16.4a_hom_XX_rec
20140211_RIC8B_E18.5_16.4a_hom_NA_rec

FIGURE 6: Identification section. A red circle has been added for this image to highlight where a non-canonical name has been used. A blue circle identifies where the name has been edited and the yellow circle identifies the change which is made after “update” is pressed. The naming of all future files and folders is also updated when the “update” button is pressed. NOTE: See pdf version of user guide for higher resolution image.

μCT Specific Fields

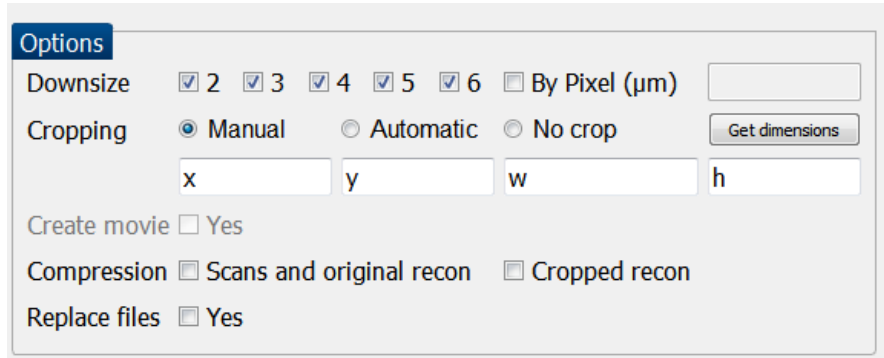
These fields are specific to using the μCT modality. The fields are autocompleted after the input folder is selected. If a file is not found the text “Not found” is shown. In this situation the user can select the folder manually by clicking on the corresponding button on the left.

μCT only	
Pick recon log	IC8B_E18.5_16.4c_wt_rec.log
Pick SPR	E18.5_16.4c_wt_rec_spr.bmp
Scan folder	Not found

FIGURE 7: μCT only section. The Scan folder was not found so the user would have to get the location of the scan folder manually in order to perform any functionality on the scans.

Options

The options section allows the user to choose the downsize, cropping and compression options.



The screenshot shows a software window titled "Options". It contains several sections: "Downsize" with checkboxes for factors 2, 3, 4, 5, 6, and a checkbox for "By Pixel (µm)" with an adjacent text box; "Cropping" with radio buttons for "Manual", "Automatic", and "No crop", and a "Get dimensions" button; four text boxes labeled "x", "y", "w", and "h" for manual cropping dimensions; "Create movie" with a "Yes" checkbox; "Compression" with checkboxes for "Scans and original recon" and "Cropped recon"; and "Replace files" with a "Yes" checkbox.

FIGURE 8: Options section.

The **downsizing** of the image is carried out in ImageJ and is performed after the cropping. The downsizing can be performed by a factor of 2 to 6 or alternatively a desired pixel size can be specified for the downsized image. No interpolation is used but the images are averaged when downscaling by a factor between 2 to 6. When downscaling by a desired pixel size, Bicubic interpolation is used. An individual tiff stack is created for each downsized selection.

The **cropping** of images is performed prior to the downsizing can either be performed manually, automatically or not at all. If the user chooses the manual option they can either enter the dimensions into the x, y, width and height text boxes or press the “Get dimensions” button.

The “**Get dimensions**” button will open up a “**z projection**” of the original reconstructed images. This effectively gets the brightest point through the stack. The user can then select a region of this image to use for the cropping dimensions. Be aware, this can take approximately **30 minutes for large files**.

The **compression** options allow the user to archive and compress the scan, original recon, and cropped folder. The archiving is carried out by “tar” and the compression is carried out with “Bzip”. If a Windows system is being used we recommend using **7-zip** to access the archived and compressed folders. It should be noted here that the folder being compressed will **not** be deleted.

When the “**replace files**” option is checked the warning flags suppressed when the output folder already exists. The result of this means that items in the output folder might be replaced if they have the same name as the new files being created.

The “**create movie**” option is not currently available.

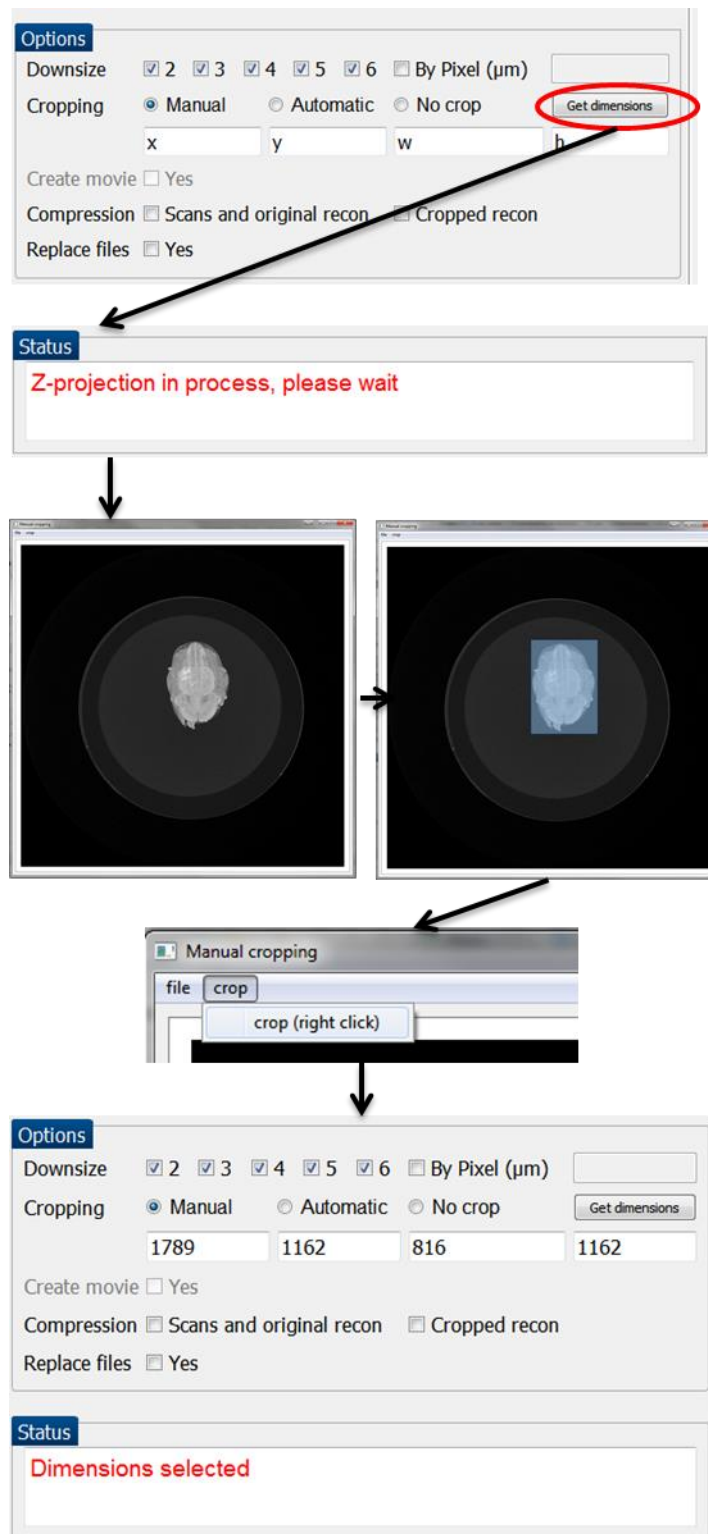


FIGURE 9: “Get dimensions” Steps to get dimensions manually from a Z-projection image. The generation of the Z-projection may take approximately 30 minutes for files >30gb over a network. NOTE: See pdf version of manual for higher resolution image.

Status section

This section just updates the user about what HARP is doing when the “Get Dimensions” button is clicked.



FIGURE 10: Status section.

Add to Processing List

When all the parameters have been assigned the “Add to Processing List” should be clicked. This will generate a folder at the specified location for the processing to be performed. Multiple folders of reconstructed images can be lined up to be processed.

The same output folder cannot be used more than once in the processing list.

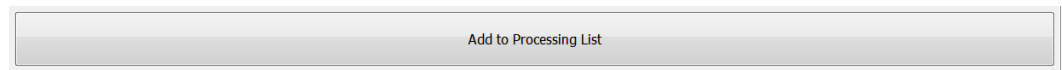


FIGURE 11: Add to Processing List button

Processing Tab

The “**Processing**” tab displays a list of the recons which are to be processed. At the top of the tab there are three options “Start”, “Stop” and “Add More”. Below these buttons there is a table which displays the recons to be processed. See FIGURE 2.

When the “**Start**” button is clicked processing will begin starting with the first recon in the table which has a status of “**Pending**”. The “Start” button will then become greyed out until the processing has finished.

The “**Stop**” button can be clicked at any stage following the start of processing and will stop any processing currently being performed and stop further processing. Any files which that were generated prior to pressing “stop” will still be saved.

The “**Add more**” button should be pressed when an additional recon folder is required to be processed. It simply opens up the “Parameters” tab.

The Processing list table consists of 5 Columns. The “**Input**” column displays the location of the original recon. The “**Output**” column consists of the output folder. The “**Status**” column lets the user know what processing is being performed. The “**Start Time**” and “**Finish Time**” record how long the processing took.

To **delete** or **remove a row** the user should click on the row to be removed and the “delete” key on the keyboard should be pressed. This will remove the selected recon folder from being processed. If the recon folder has already been processed this will just remove the information from the table but **will not remove any files that have been created**.

Output Summary

The output from HARP will be stored in the specified output folder. It will consist of a **“Metadata”**, **“Cropped”** and **“Scaled stacks”** sub directory (unless no cropping or scaling was performed). If the “compress cropped recon” option was selected then an additional archived and Bzipped file will be present in the folder.

See below for list of folder contents

<i>Folder</i>	<i>Folder contents</i>
Output/ e.g 20140408_RCAS_17_18.4e_wt_rec/	Metadata/ Cropped/ Scaled stacks/ Compressed cropped folder: Cropped_[date]_[gene]_[stage]_[litter]_[zygosity]_[sex]_rec ..tar.bzip
Cropped/	Cropped recon files Additional files from original recon
Scaled stacks/	A “tiff” stack for each scaling type. The name will include the scaling factor and the new pixel size. e.g 20140408_RCAS_17_18.4e_wt_rec_scaled_x2_pixel_8.9 626
Metadata/	config4user.txt configobject max_intensity_z.tif pid.log scale.log session.log

The **“Metadata”** folder contains information used for processing, tracking and debugging.

The **configobject.txt** file is a python “pickle” file which includes the parameter settings. The **config4user.txt** file is an easier to read version of this pickle file.

The **max_intensity_z.tif** file is the z-projection used for “Get dimensions”.

The **session.log** file is a record of what processing occurred.

The **scale.log** file is a record of what occurred in the scaling performed by ImageJ.

The **pid.log** file identifies what processing threads are being currently used. This is useful for debugging.

Error and Warning Messages

A number of warnings and error messages will flag up if the parameters or naming conventions have not been met. Below details these warnings and gives the user further instructions.

Warning/Error Message	Further instructions
Warning: Name ID is not in the correct format	Either proceed with non-canonical name or edit as described in "Identification" section
Warning: input directory not defined	The user needs to select an input directory. See "Input and Output"
Warning: input folder does not exist	The input folder should be a single folder consisting of files generated from a reconstruction. See "Input and Output"
Warning: input folder is empty	The input folder should be a single folder consisting of files generated from a reconstruction. See "Input and Output"
Warning: output directory not defined	The output folder is required to store any processed images. See "Input and Output"
Warning: Scan folder not defined	Scan folder needs to be defined if the compression of the scanned folder is required. See "Options".
Warning: Scan folder does not exist	Scan folder needs to be defined if the compression of the scanned folder is required. See "Options".
Warning: User has not specified a new pixel size value	If the user wants to scale the processed image to a certain pixel size, the pixel size has to be defined. See "Options".
Warning: User defined pixel is not a numerical value	The pixel size used for scaling by pixel needs to be numerical. See "Options".
Warning: Pixel size could not be obtained from original recon log. Scaling "By Pixel (um)" is not possible	The pixel size is determined from the original recon log file. See "Options".
Warning: Cropping dimensions have not been defined	If cropping using "manual" setting user supplied or "Ge dimensions" supplied dimensions are

	required.. See “Options”.
Warning: Output folder is already on the processing list'	To avoid unwanted overwriting of files. An output folder can only be used once in the processing list. If the entry is removed from the processing list, the output folder can be added. See “Processing Tab”
Warning: Can't delete a row that is currently being processed. Select "Stop", then remove	The user can only delete items off the processing list which have either finished processing or have not started processing. See “Processing Tab”
Cropping Error; see session log file	An error has occurred while performing cropping the Session log will show more details of this.
Warning: Unexpected error	Sorry! As this was unexpected we are not sure what went wrong here without having a look. Please let a member of SIG know about the problem.

Appendix

Required python modules

argparse
autocrop
collections
ConfigClass
cPickle
crop
datetime
fnmatch
logging
math
matplotlib
multiprocessing
numpy
operator
os
pickle
PIL
pprint
psutil
PyQt4
re
shutil
signal
subprocess
sys
tarfile
tempfile
time

uuid

Contact details

Henrik Westerberg: h.westerberg@har.mrc.ac.uk

Neil Horner: n.horner@har.mrc.ac.uk

Tom Lawson: t.lawson@har.mrc.ac.uk