MRC HARWELL

Systems Imaging Group

HARP
User Manual

Table of Contents

Section

1

Introduction

To his 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 μCT and OPT scans and reconstructions.

The intended end user is a researcher working on the IMPC embryo phenotyping pipeline.

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

James Brown: james.brown@har.mrc.ac.uk

Getting Started

Windows

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

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:\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 in order for HARP to run (see Appendix). 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 harp.py

Mac

HARP can be run on a Mac in the same way as on Linux.

Known issues

On some operating systems (primarily Debian), HARP's user interface may "hang" during processing. This is a known issue that the developers are attempting to resolve. For now, we suggest queueing up all of the jobs to be processed prior to clicking "Start", as the user interface may become slow or unresponsive when attempting to add new jobs. If you experience this issue or any other problems, please contact the developers.

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 so it will look different if HARP is running in Linux or Mac.

Overview

Change view and access help files

Select modality

Tab select

Select input and output

Recon info

ID and nomenclature check

Processing options

Select additional files/folders

OPT channels and options

Status updates (e.g. when "Get dimensions" is clicked)
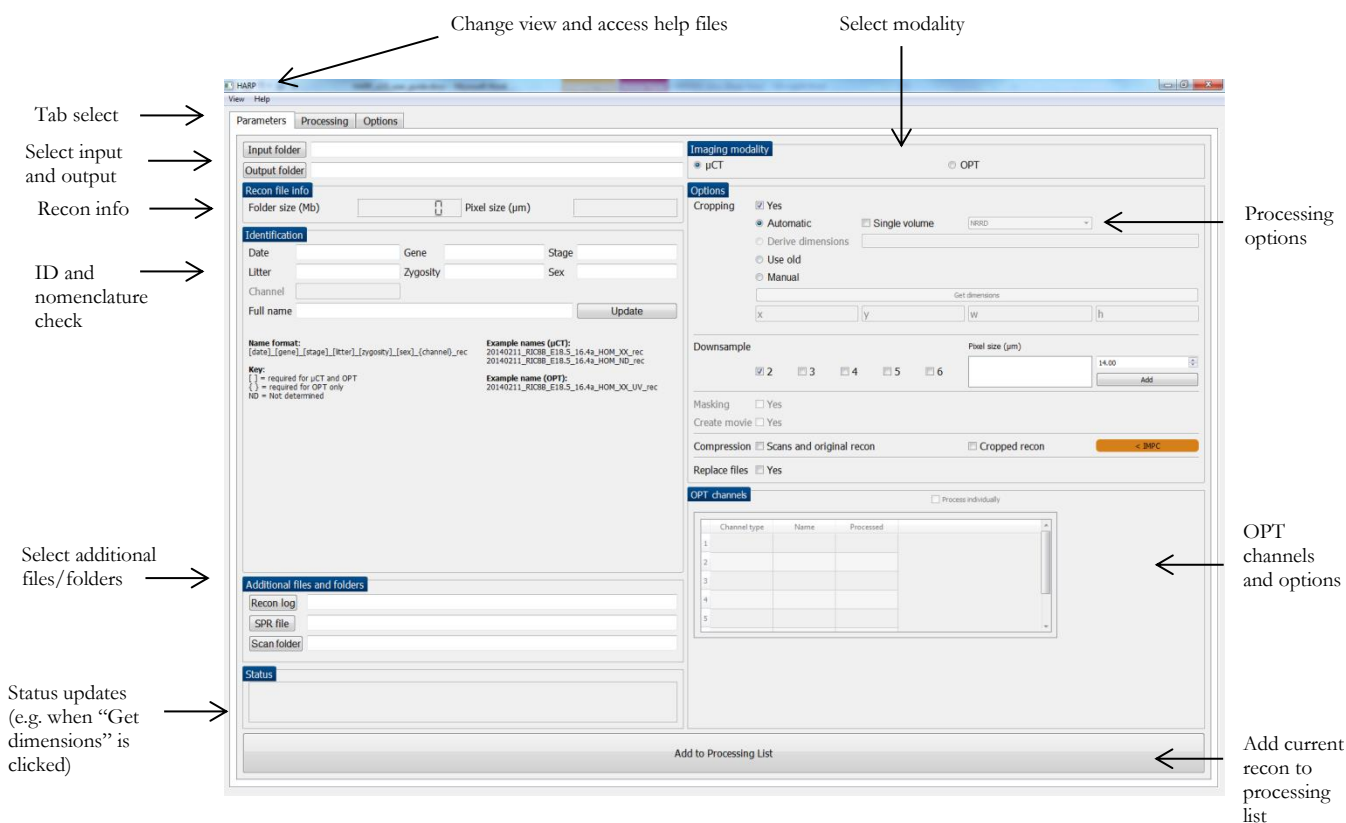
Add current recon to processing list

FIGURE 1: Starting view of HARP. The current selected tab "Parameters". Any functionality not currently available is greyed out.
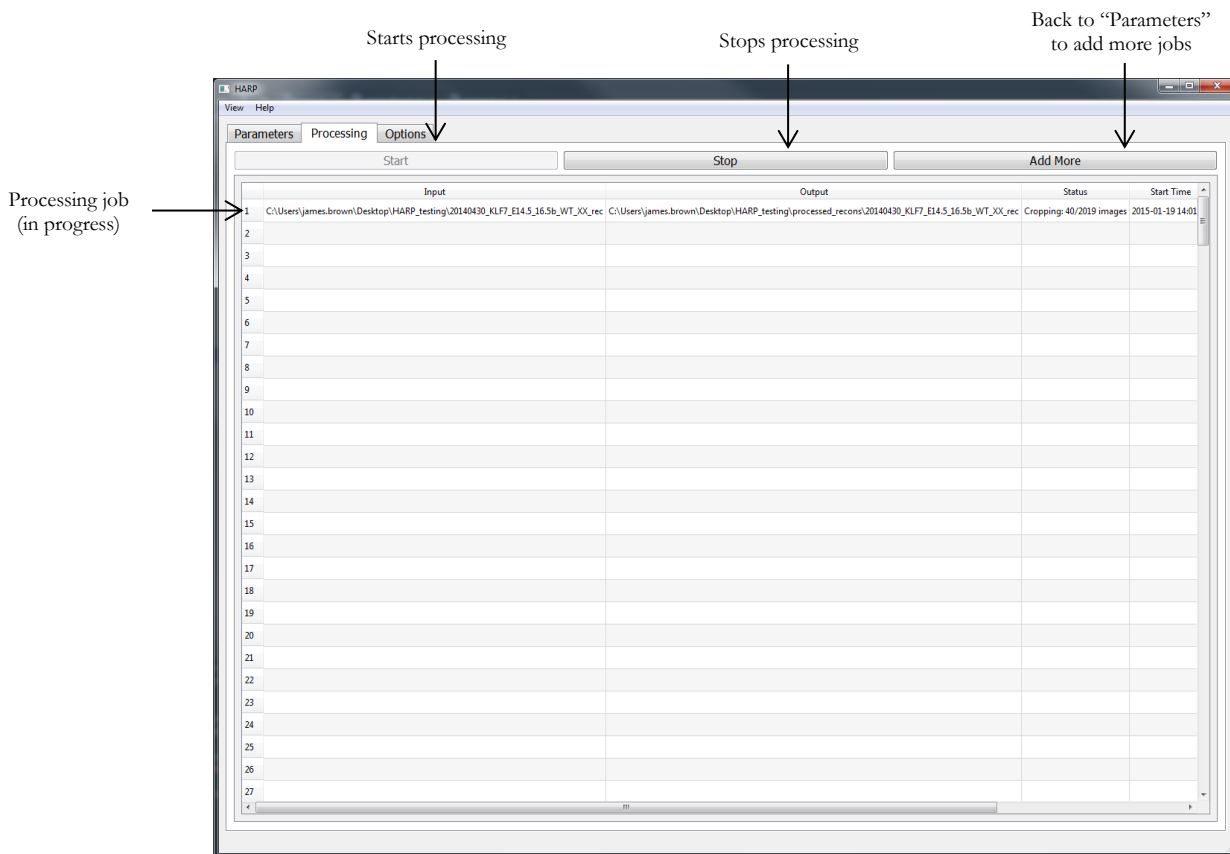
FIGURE 2: Processing view of HARP, shown when the "Processing" tab is selected and whenever a user successfully adds a recon folder to the processing list.
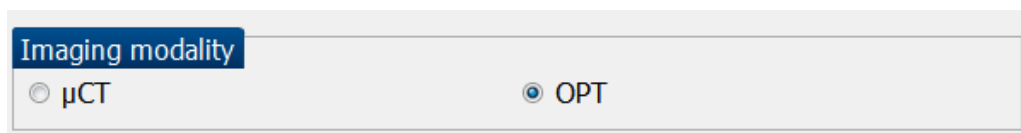
Input and output select

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

After a folder has been selected, HARP attempts to auto-populate a number of fields. This includes suggesting an output folder, identifying if OPT or μCT, calculating the recon folder size, identifying the pixel size, checking the identification nomenclature, and additional files and folders (log file, SPR file, 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 directory. Alternatively the output folder can be assigned manually by clicking on the **"Output folder"** button.

Imaging Modality Selection

HARP can process μCT or OPT data. HARP will try and identify this automatically but it can be manually selected, if it fails to do so.
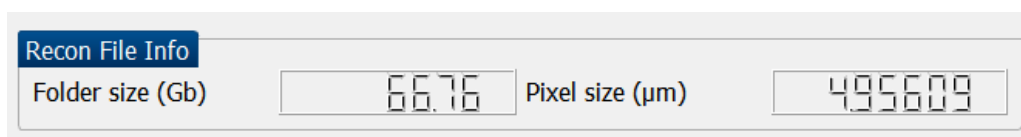

FIGURE 4: Imaging modality selection.

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.

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 date, gene, stage, litter, zygosity and sex separated by an underscore. If one of the identifiers is not known the letters NA or ND should take its place. OPT naming should also include the channel. See below for naming conventions.

Box 3: μCT Nomenclature

Name format:
[date]_[gene]_[stage]_[litter]_[zygosity]_[sex]_rec
Example names:
20140211_RIC8B_E18.5_16.4a_hom_XX_rec
20140211_RIC8B_E18.5_16.4a_hom_ND_rec

If the wrong name format has been used a warning box will pop up. The user can carry on with this naming format or can change the name for all processed data.

Edit the recon name here (if required)

Then click to update
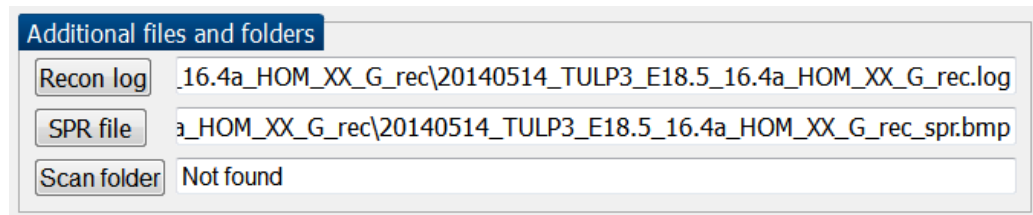
FIGURE 5: Identification section. An orange circle has been added to highlight where a non-canonical name has been used. A blue circle identifies where the name has been edited and the black circle identifies the change which is made after "update" is pressed. The naming of all future files and folders is also updated when the "update" button is pressed.

Additional files and folders

HARP automatically locates the recon log, SPR file and the scan folder. The recon log file is required if the recon is to be downsized to an arbitrary pixel resolution.

The scan folder is required if the user wants to compress the scan folder. If a folder can't be automatically located the text **"Not found"** will be shown. These be manually specified, or ignored if not required by the processing.
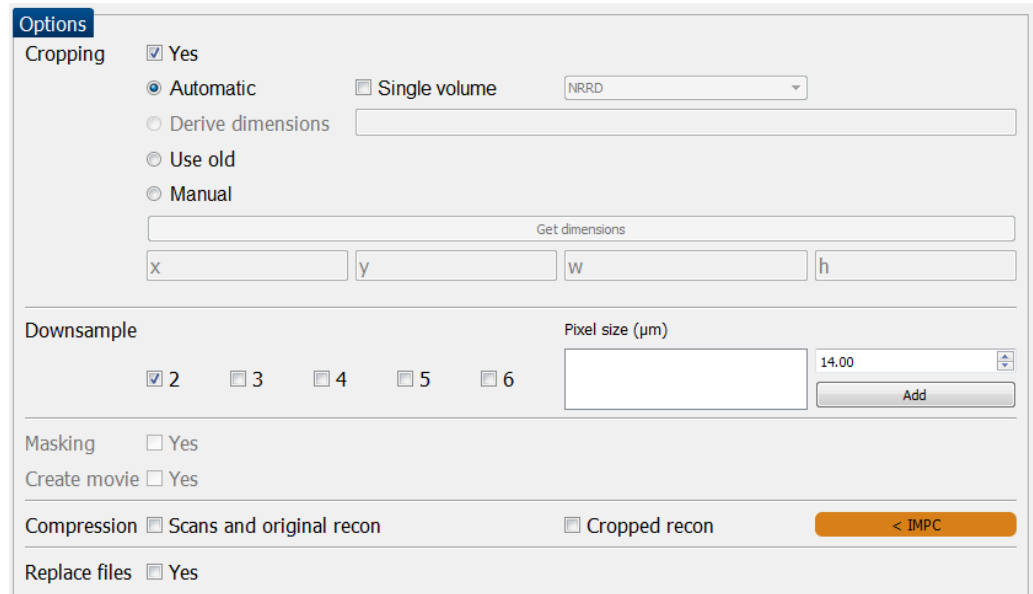
FIGURE 7: Additional files and folders section.

Options

User can decide on cropping, downsizing and compression options.

FIGURE 8: Options section.

The **cropping** of images can be deselected by un-ticking the "Yes" box. When this box is not ticked HARP will downsize the original reconstruction image.

When the **"Yes"** box is ticked the following options are available for cropping:

- **Automatic**: The cropping dimensions or crop-box will be determined automatically in the X and Y dimensions. The Z dimension will not be affected.

  Derive dimensions (OPT only): This option provides additional functionality that for most instances is not required.

  The option allows the crop dimensions to be derived from a previously processed channel. In the OPT channels section the user should double-click the appropriate channel, or select it and press "Return"

- **Use old:** If HARP has previously been used to perform the cropping of a recon, the cropped image can be used to perform additional processes such as downsizing.

  e.g. The user has already cropped and downsized by 3 and 4 but later realized they needed to downsize by 5 and 6 as well. The user would setup HARP as before, but also select downsize by 5 and 6 and select "Use old". This way the cropping step is not repeated unnecessarily.

- **Manual:** In cases where the automatic crop is not capable of determining the crop dimensions, the user can either enter the dimensions into the **x, y, width** and **height** text boxes or press the **"Get dimensions"** button.

  The **"Get dimensions"** button provides a Z-projected image of the stack, (brightest pixels through the stack), and allows the user to select a cropping region by dragging a box around the region.

- **Single volume:** After cropping, an unscaled file can be saved as either a TIFF stack or NRRD file by selecting this option.

The **downsizing** of the image is performed after the cropping. The downsizing can be performed by a factor of 2 to 6, or alternatively a series of pixel sizes 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, "area" interpolation is used (OpenCV 2.0). An individual NRRD (Nearly Raster Raw Data) file is created for each downsized selection.

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. See Appendix for 7-Zip instructions.

When the **"replace files"** option is checked the warning flags are 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"** and **"mask"** options are not currently available.

1. Select "Manual" and click "Get dimensions"

2. Wait for Z-projection to finish (see status updates)

3. Click and drag box around region to be cropped (adjust by moving corners)
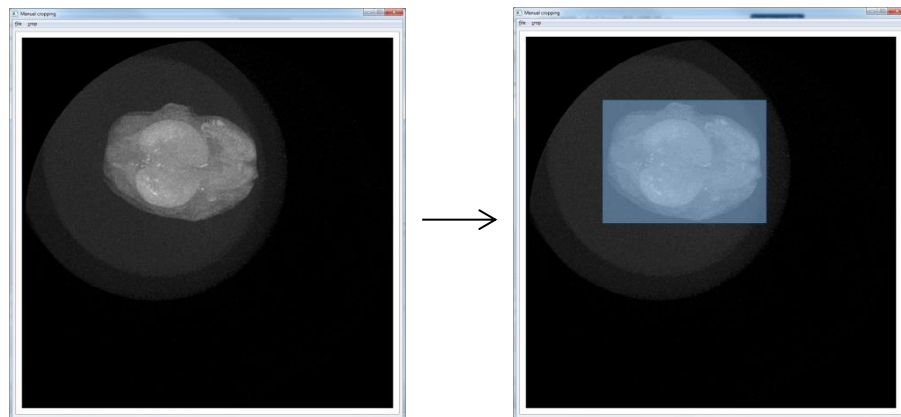
4. Click "crop (right click)" in the crop menu or right click anywhere in the window to confirm the selected crop box
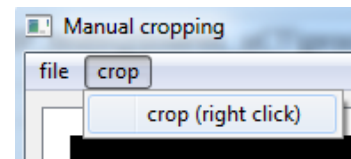
5. The dimensions fields and status text will be updated upon confirmation

FIGURE 9: Step-by-step guide to get dimensions manually from a Z-projection image.

OPT channels

FIGURE 9: OPT channels section.

This section displays all the OPT channels HARP could identify for the selected recon folder. If the folder is not in the correct naming format the OPT channels will not be displayed here. By default all OPT channels will be added to the processing list together and use the channel which is highlighted in red to determine the crop box dimensions.

For the example in the figure there are 4 channels: UV, white, red and green light. The white light is highlighted as red. This means the crop box dimensions will be determined using the white light channel. All other channels will derive their cropping dimensions from this channel. The user can double click on any channel to change which channel should be used to determine the crop box dimensions though.

Alternatively if the **"Process Individually"** checkbox is selected the channels can be processed individually. The user can then choose the cropping method used for each channel.

The **"Processed"** column details whether the channel is either already processed or on the processing list.

Status section

This section updates the user about what HARP is currently doing when the "Get Dimensions" button is clicked, and whether the chosen dimensions were confirmed.

FIGURE 10: Status section.

Add to Processing List

When all the parameters have been assigned, click the **"Add to Processing List"** button. 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.
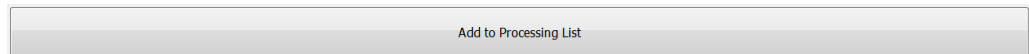
Add to Processing List

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"** or **"Processing cancelled!"** Any finished jobs will not be restarted. The **"Start"** button will then become greyed out until the processing has finished, or if the **"Stop"** button is pressed.

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. Any cancelled jobs will be restarted upon clicking the **"Start"** button again (unless removed – see below).

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 five columns. The **"Input"** column displays the location of the original recon. The **"Output"** column displays the specified 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 press the "delete" key on the keyboard. This will remove the selected recon folder from the processing list. If the recon folder had 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 **"Cropped recon"** compression option was selected then an additional archived and Bzipped file will be present in the folder.

See below for list of folder contents:

Folder	Folder contents
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 "NRRD" file 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_2_pixel_8.9626.nrrd
Metadata/	config4user.txt configobject max_intensity_z.png pid.log X_scale.log e.g. 2_scale.log session.log

See below for further details about the metadata:

The **session.log** file is a general record of what processing was performed.

The **X_scale.log** file is a record of what occurred in the scaling.

The **configobject.txt** is a Python "pickle" object used to save the parameters HARP uses. The **config4user.log** is user friendly human-readable version of the pickle file.

The **cropbox.txt** file is Python pickle file of the cropping dimensions used.

The **max_intensity_z.png** is the image used for manual cropping using the "Get dimensions" option.

Appendix

7-Zip instructions

7-Zip is a free and open source compression and archiving tool for Windows.

Download from: http://www.7-zip.org/download.html

To un-archive and unzip:
7-Zip should be accessible from right clicking on the document. The user can navigate to "Extract here". This will unzip the file.
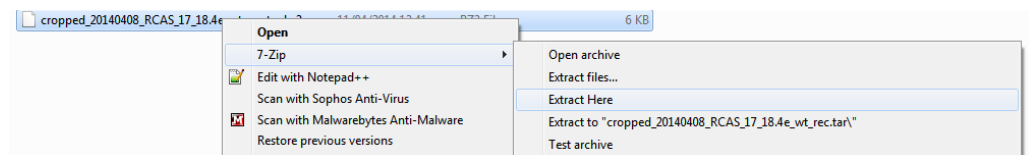

FIGURE 12: 7-Zip first step.

The user should then right click and "Extract to …" to un-archive the folder.
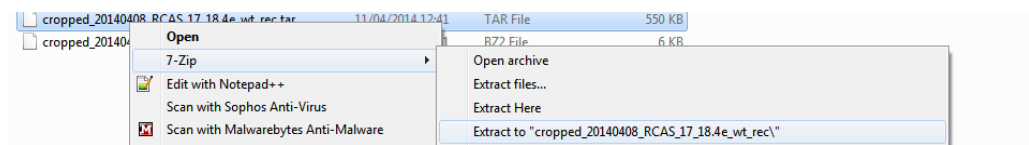

FIGURE 13: 7-Zip second step.

Alternatively 7-Zip file manager can be opened up and the steps shown previously can be carried out inside the browser.
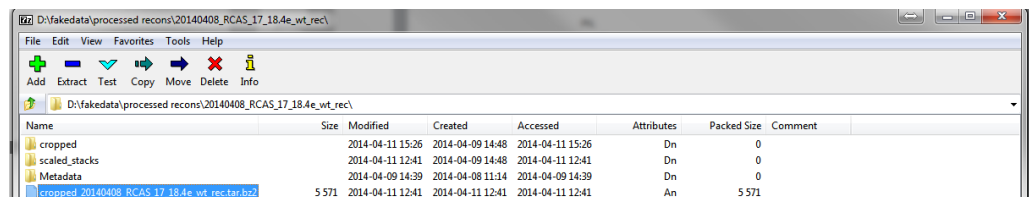

FIGURE 14: 7-Zip file manager

Required external Python modules

If you wish to run HARP as a Python script (or build from source), you will need to install a number of Python modules. The source code can be found on Github (https://github.com/Tomnl/HARP) and downloaded either as a **.zip** or via the command:

> **Box 3: Clone HARP github repository**
>
> git clone https://github.com/Tomnl/HARP.git

Included within the source code is a setup script which may be executed in order to install the necessary modules:

> **Box 4: Run setup script**
>
> python setup_HARP.py

Note that the 'easy_install' utility must be installed in order for this script to work. Unfortunately Qt4 and its Python bindings (PyQt4) cannot be installed this way and must be installed separately.