

# **DDSM Utility v2.0**

**By**

**Anmol Sharma**

**Undergraduate Engineering Student**

**DAVIET Jalandhar**

**Punjab, India**

**Email: [anmol.sharma293@gmail.com](mailto:anmol.sharma293@gmail.com)**

# Introduction

---

This is a tutorial for the DDSM Utility v1.0 written by me (Anmol Sharma, Undergrad Engineering Student at DAVIET Jalandhar) to greatly simplify downloading, converting, viewing and extracting annotations from the Digital Database for Screening Mammography (DDSM) database available here:  
<http://marathon.csee.usf.edu/Mammography/Database.html>

DDSM is a very famous mammogram database which researchers around the world use to test their mass detection and mass classification algorithms so as to ensure that the results are comparable to other authors or systems. However, an open source utility or tool doesn't seem to exist to easily and efficiently download the database. The website where the database is hosted has not been updated from the last 15 (almost 16) years, and hence the deprecated and obsolete software that comes with it is essentially, unusable. None of the authors using the database has released code that can process DDSM images, with their research.

The utility is capable of automatically parsing ICS files which contains the LJPEG image sizes to facilitate LJPEG to LJPEG1 decompression, and further conversion into other known formats like PNG. I decided to write this utility when I was looking for ways to download and convert DDSM dataset, only to find out that the images are in a very old LJPEG format. The only utility available on the DDSM website ([here](#), written for very old SunOS 5.6) was to convert it into LJPEG1 (another old, obsolete and unusable format), and that too was highly labor intensive because it required the user to manually run the script for every image he wanted to convert. Even if the user did that, converting LJPEG1 file to PNG was a nightmare, if not impossible.

Another utility available was written by Dr. Chris Rose of University of Manchester ([here](#)). It required the user to input names of IMAGE files manually, which then his script used to download the DDSM image in PNG format. This was again, ridiculously labor intensive, and didn't make sense if the user wants to download, say, 1000 images. The source code was not supplied, so making changes to the internal working was not possible. Moreover, if the user wanted to download the annotation files along with the image, he has to repeat the steps all over again. I wasn't getting paid by the hour, so this was downright ridiculous for me.

So I decided to write my own little utility to get the job done. However it was easier said than done. Eventually I figured out that the solution will have to span two operating systems (Linux and Windows), and two languages (C++ and MATLAB). This was due to the fact that I decided to reuse a small jpeg utility supplied by DDSM which only ran on UNIX based platforms (originally SunOS 5.6), and partially due to simplicity of MATLAB in reading, displaying and writing image files as compared to C++.

# Contents

---

My utility contains the following files:

<b>openDDSMJPEGAndConvertToPNG.m</b> – Main MATLAB script.
<b>ConvertDDSMImageToRaw.m</b> – Supporting MATLAB script to open the compressed JPEG files and then convert the files into PNG or any other user defined format. Uses the utility jpeg and dds2raw.exe files written by Dr. Chris.
<b>openDDSMpngWithOverlay.m</b> – MATLAB script to open the converted PNG files, read their corresponding OVERLAY files, read their annotations and display the boundary details by superimposing on the mammogram for simple viewing.
<b>readBoundary.m</b> – Supporting MATLAB script to read OVERLAY files (Original author Dr. Jayasree Chakraborty, Memorial Sloan Kettering Cancer Center, New York.)
<b>Jpeg.exe</b> - Windows executable file to convert JPEG to JPEG1 (original author Dr. Chris Rose, University of Manchester)
<b>ddsmraw2pnm.exe</b> - - Windows executable file to convert JPEG1 to PNG or any other format (original author Dr. Chris Rose, University of Manchester)

# Features of the DDSM Utility

---

- > Easily download multiple case files automatically.
- > Normalize LIPEG images according to their scanners automatically! No manual normalization needed.
- > Automatically convert LIPEG DDSM images into PNG, JPEG, TIF, GIF and other formats.
- > Read boundary information of each of the converted images and display them.

## Initial Requirements

---

- A Windows based machine with MATLAB 8+ installed.
- WinSCP
- Cygwin
- MATLAB

Tested on:

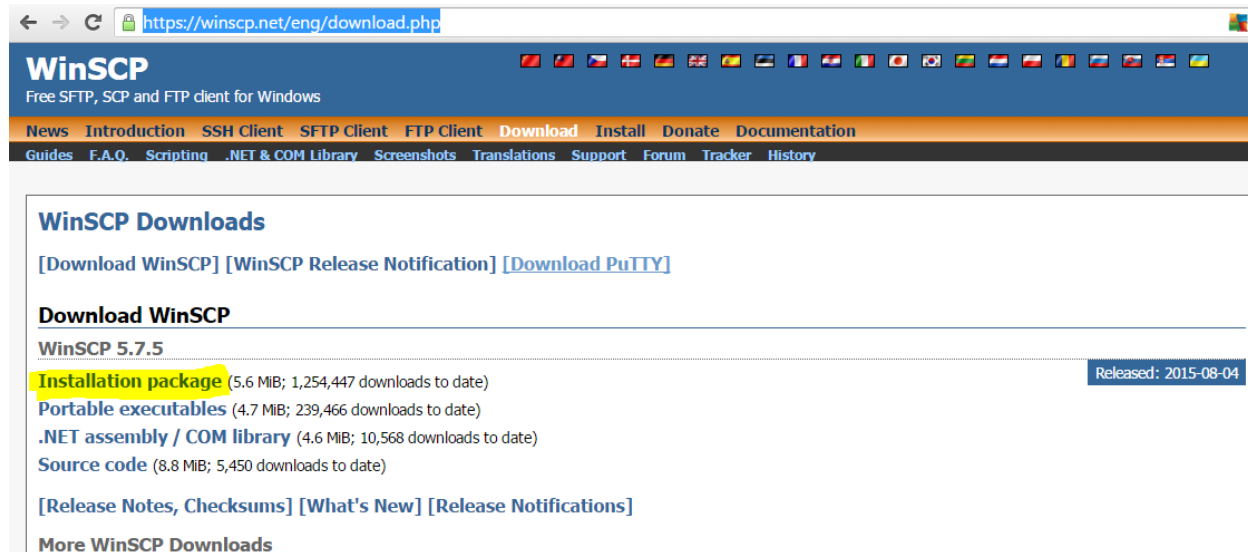
- Dell Inspiron 15 3537
- Intel i5 4200U
- 4GB RAM
- Cygwin 32 bit with ImageMagick and Ruby Interpreter.
- Windows 10 with MATLAB R2015b
- DDSM Software by Dr. Chris Rose.

# Steps

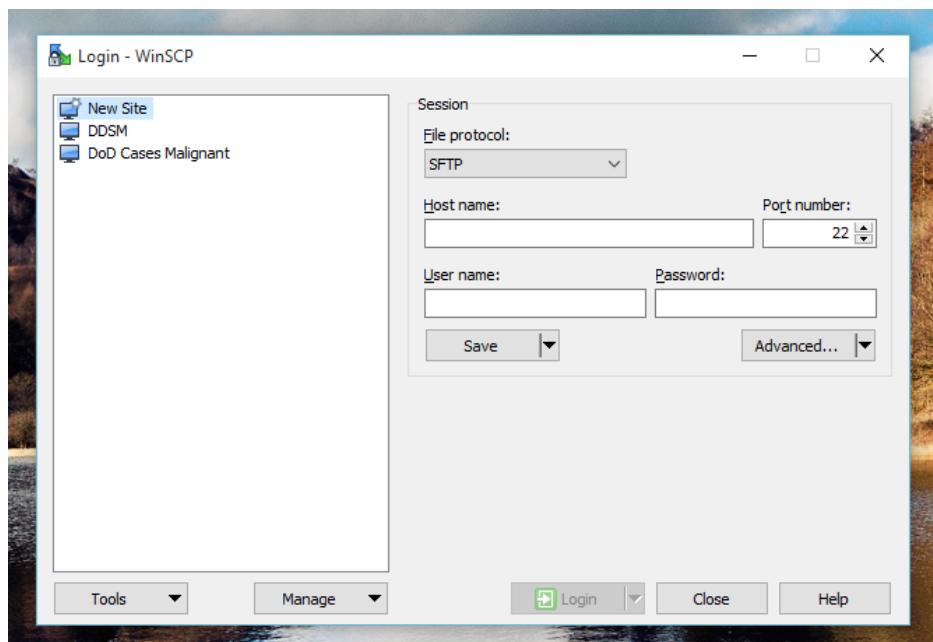
Follow the instructions and click on yellow highlighted regions in the figures. I suggest that you very carefully go through the steps, and not miss anything. Please also make sure that you perform each step in order.

## A. Download the required DDSM cases using the FTP link from USF.

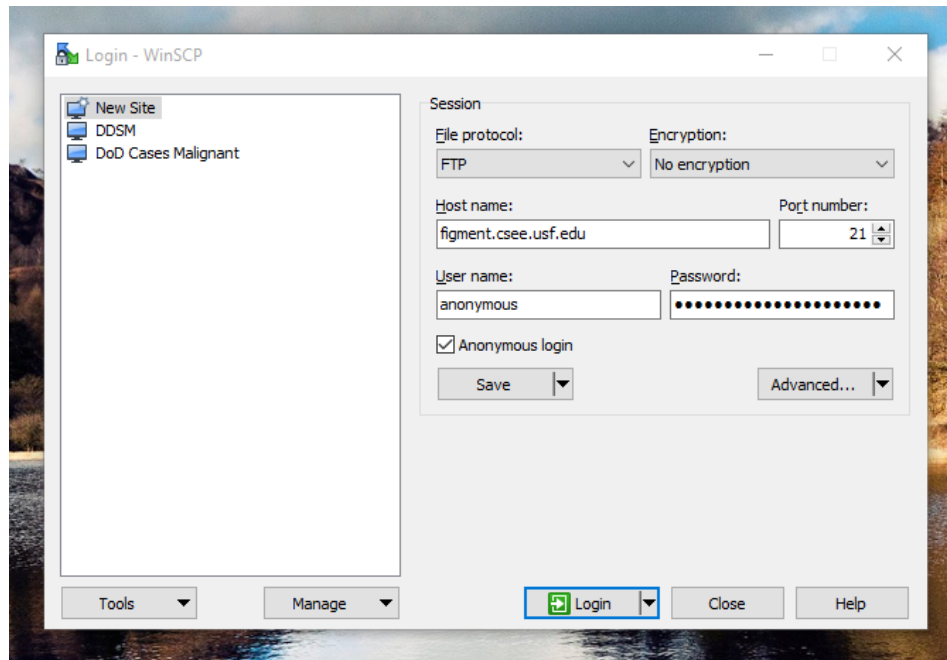
1. Download WinSCP ([here](#)).
  - a. Download the latest version as shown in figure.



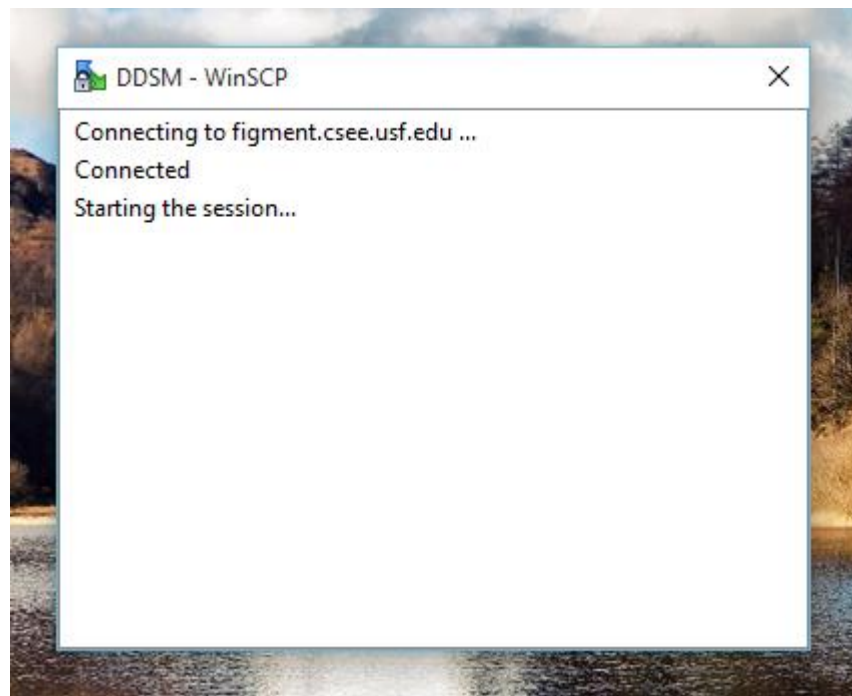
- b. Install the software, and open it.



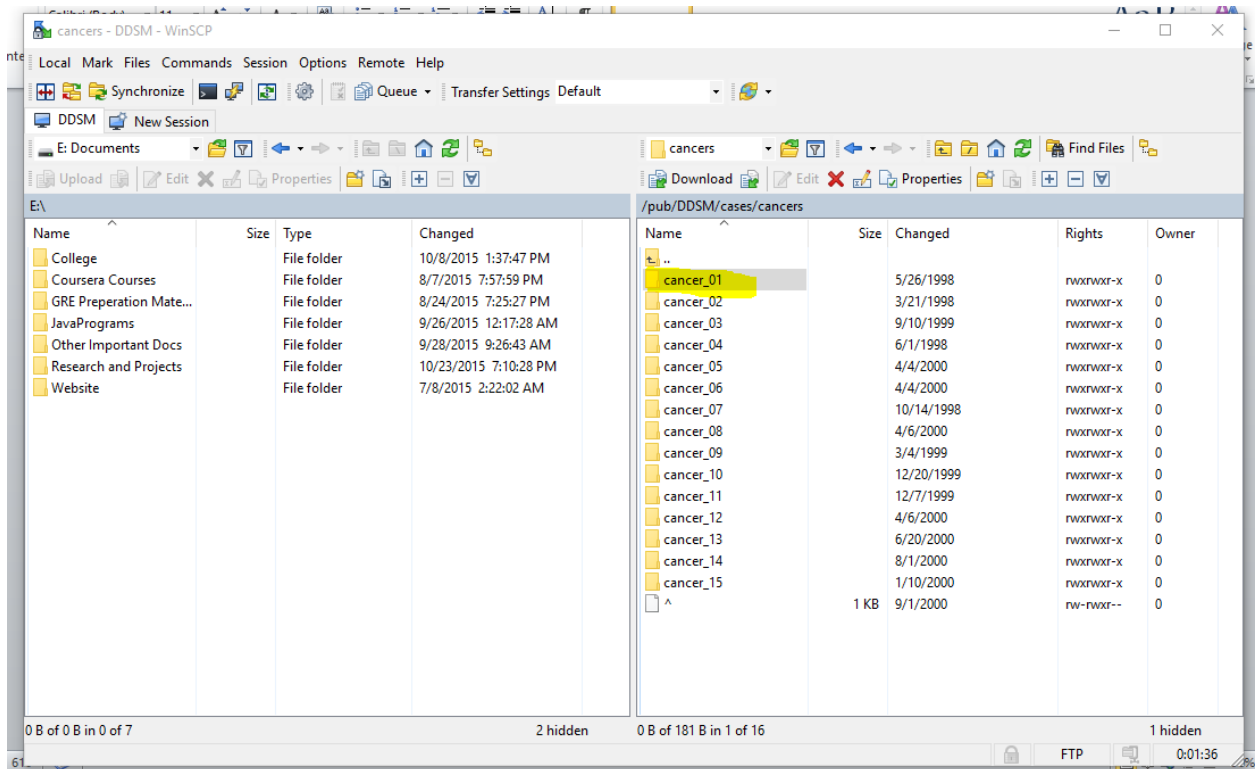
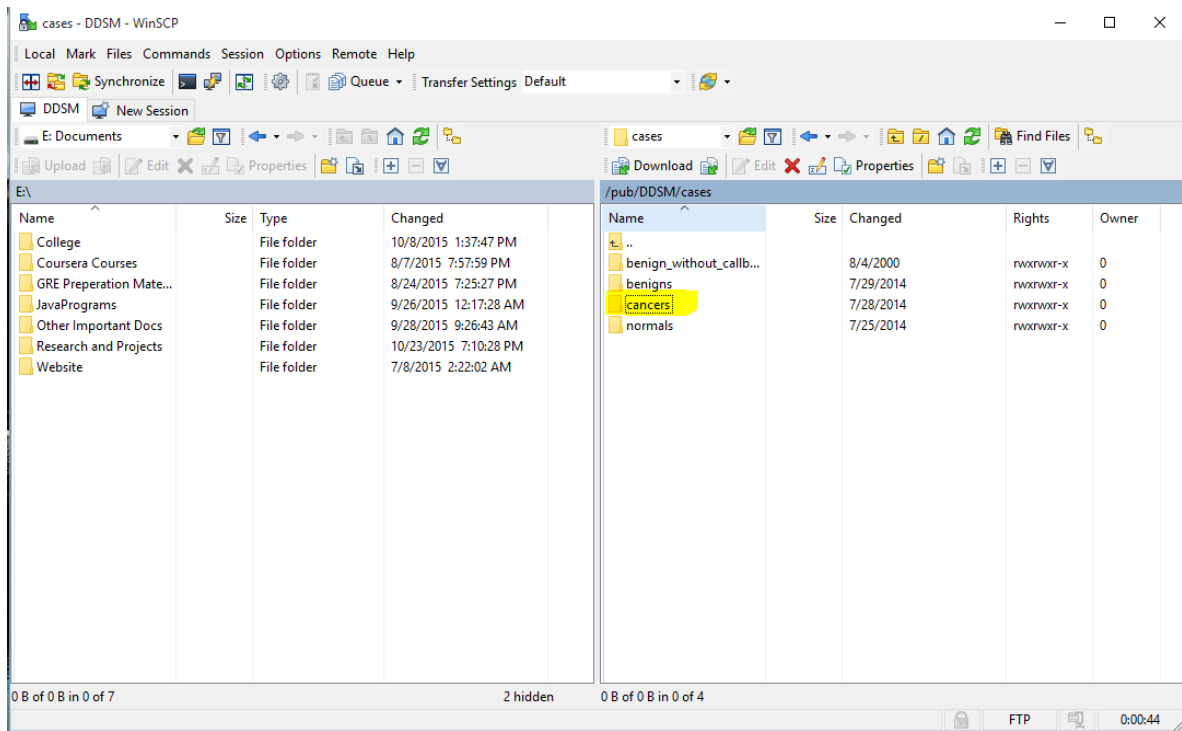
- c. WinSCP will ask you for the website and login details for the FTP server. Enter the credentials as shown below.

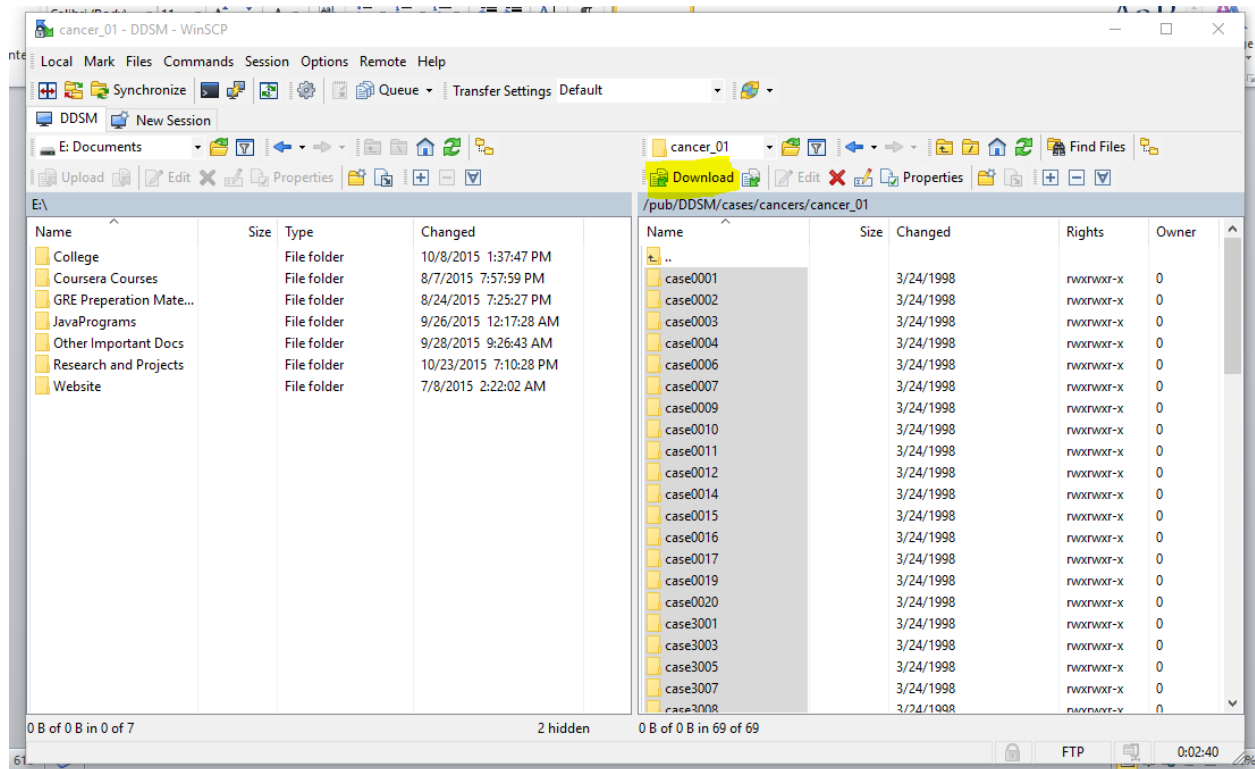


- d. WinSCP will connect to the server.



- e. Once connected, you will get the following screen. On the right, you have the FTP server parent directory, and on the left you have your own local computer's directory. Now you just need to select the files (cases) you want to download from USF's FTP server by selecting them from the right side, and then click on download button on the right to start the download to the left pane location.





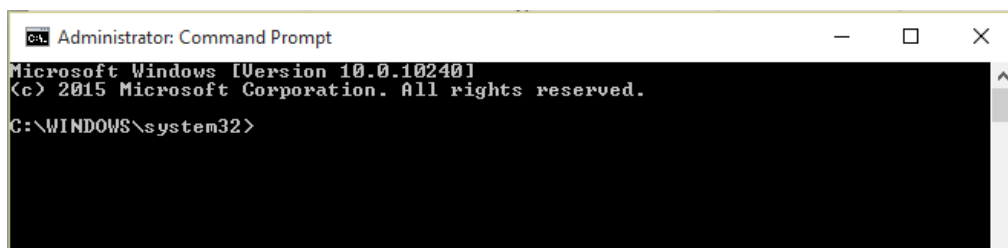
- f. I pressed CTRL+A to select all cases, and then pressed Download button as highlighted.
- g. The download will begin and you just have to wait.

## Copy all files from each of the case folder into one single folder using copy command in Windows CMD.

2. Suppose we have the following cases in a directory D:/DDSM Database Downloaded/Example/

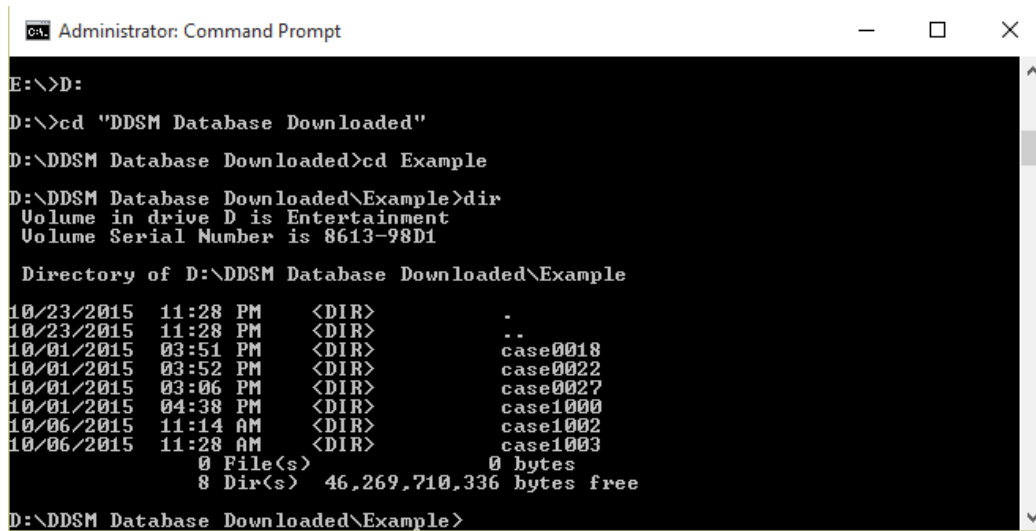
Name	Date modified	Type	Size
case0018	10/1/2015 3:51 PM	File folder	
case0022	10/1/2015 3:52 PM	File folder	
case0027	10/1/2015 3:06 PM	File folder	
case1000	10/1/2015 4:38 PM	File folder	
case1002	10/6/2015 11:14 AM	File folder	
case1003	10/6/2015 11:28 AM	File folder	

- a. Open Command Prompt with administrator privileges.





- b. Go to your directory where you have your cases folders, using **cd** command.



```
Administrator: Command Prompt
E:\>D:
D:\>cd "DDSM Database Downloaded"
D:\DDSM Database Downloaded>cd Example
D:\DDSM Database Downloaded\Example>dir
Volume in drive D is Entertainment
Volume Serial Number is 8613-98D1

Directory of D:\DDSM Database Downloaded\Example

10/23/2015  11:28 PM  <DIR>          .
10/23/2015  11:28 PM  <DIR>          ..
10/01/2015  03:51 PM  <DIR>          case0018
10/01/2015  03:52 PM  <DIR>          case0022
10/01/2015  03:06 PM  <DIR>          case0027
10/01/2015  04:38 PM  <DIR>          case1000
10/06/2015  11:14 AM  <DIR>          case1002
10/06/2015  11:28 AM  <DIR>          case1003
               0 File(s)              0 bytes
               8 Dir(s)  46,269,710,336 bytes free

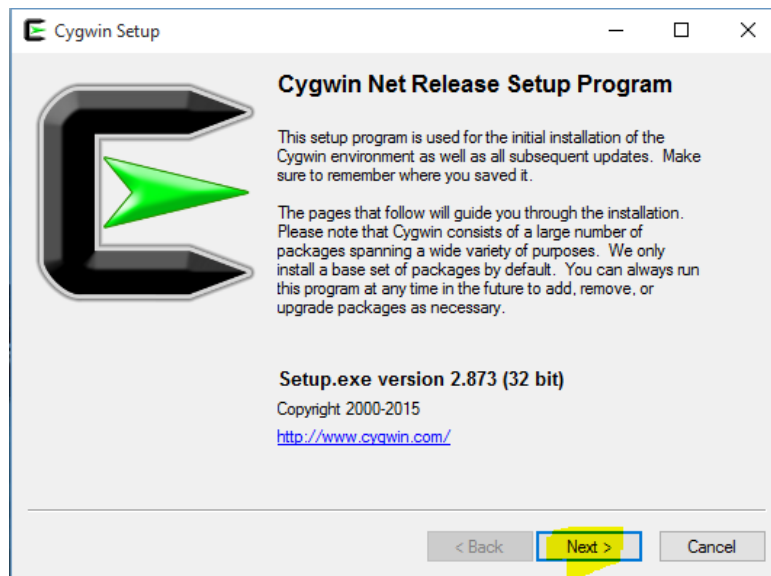
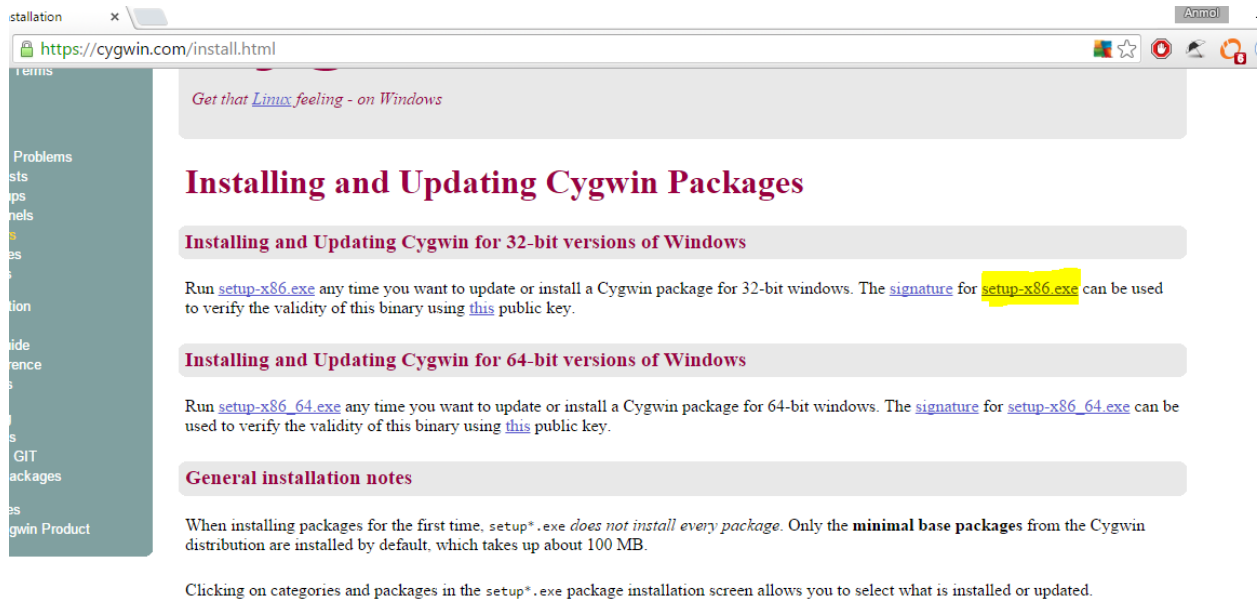
D:\DDSM Database Downloaded\Example>
```

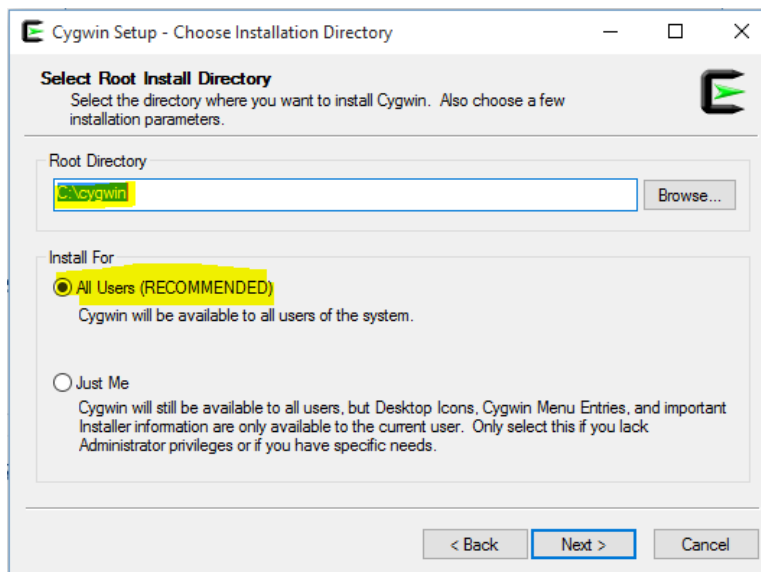
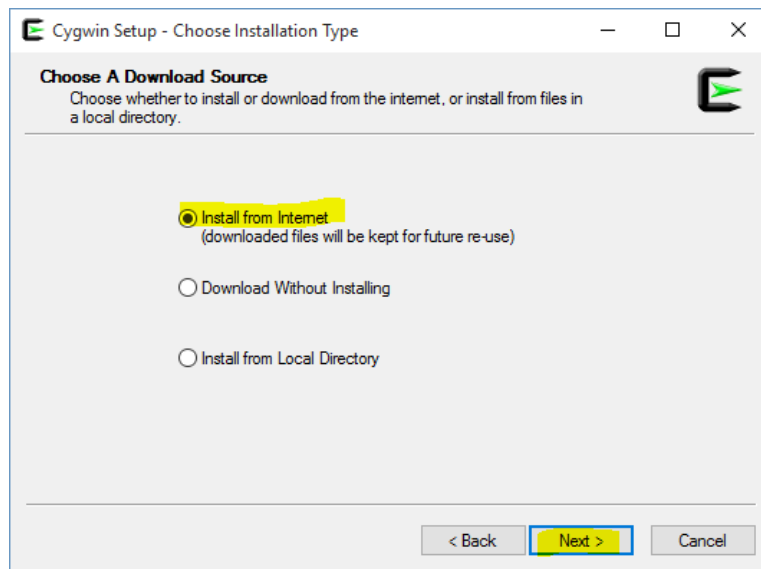
- c. Let's create a folder called "All Files" at D:/DDSM Database Downloaded/All Files where we want to copy all the files present in the case folders.
- d. Now go to your command prompt, and write the following commands one by one. Please change the source and destination of the command according to your own PC configuration. The format is:
- for /R "source" %f in (\*.extension) do copy "%f" "destination"*
- e. I used the following commands to copy all files inside case folder into All Files folder.

```
for /R "D:\DDSM Database Downloaded\Example\" %f in (*.LJPEG) do copy "%f" "D:\DDSM Database Downloaded\All Files\"
for /R "D:\DDSM Database Downloaded\Example\" %f in (*.OVERLAY) do copy "%f" "D:\DDSM Database Downloaded\All Files\"
for /R "D:\DDSM Database Downloaded\Example\" %f in (*.ics) do copy "%f" "D:\DDSM Database Downloaded\All Files\"
for /R "D:\DDSM Database Downloaded\Example\" %f in (*.16_PGM) do copy "%f" "D:\DDSM Database Downloaded\All Files\"
```

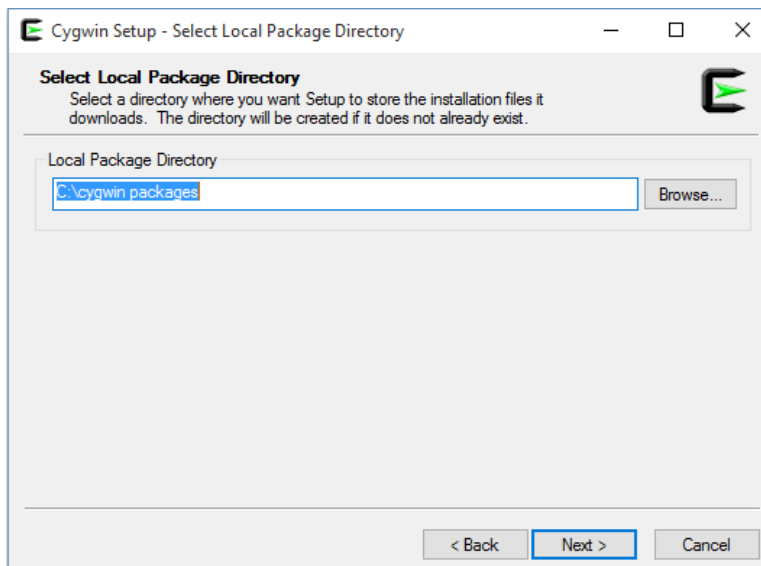
Note: Do not try to copy commands from here. The CMD terminal will reject the command due to unwanted formatting elements being included in the copied version. It is better to write it once yourself, and then press arrow keys to change the next versions.

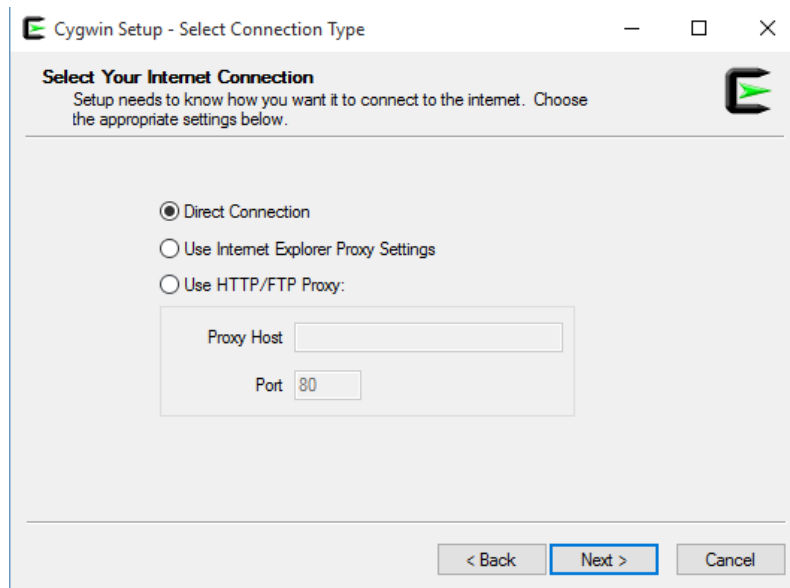
## Download and install Cygwin 32 bit on your system.



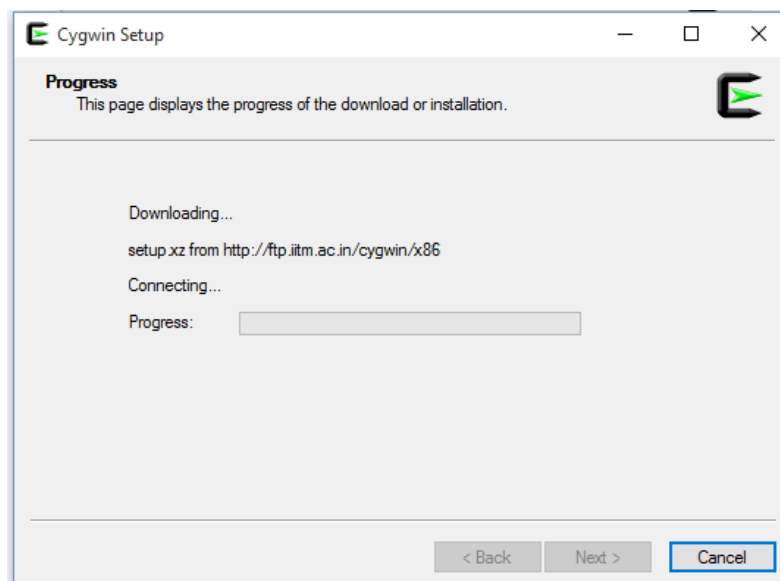
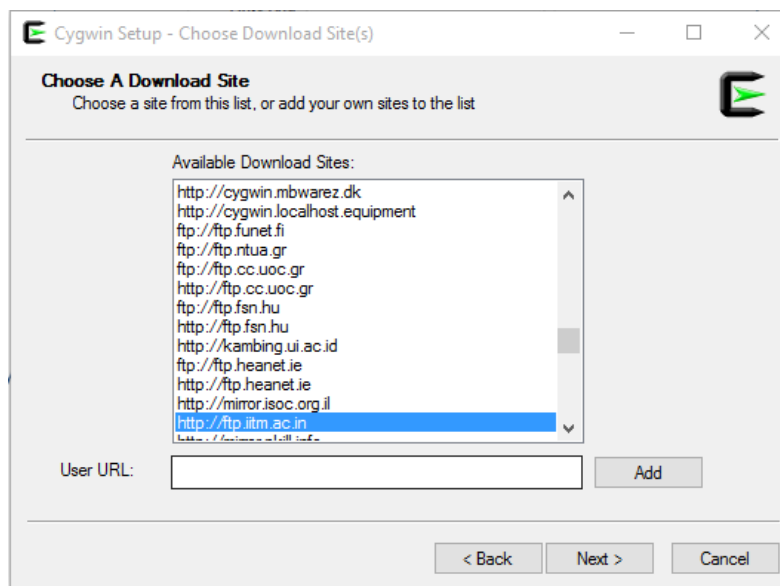


3. Create a new directory “Cygwin packages” in C:\ and use this to store your packages, as shown below.

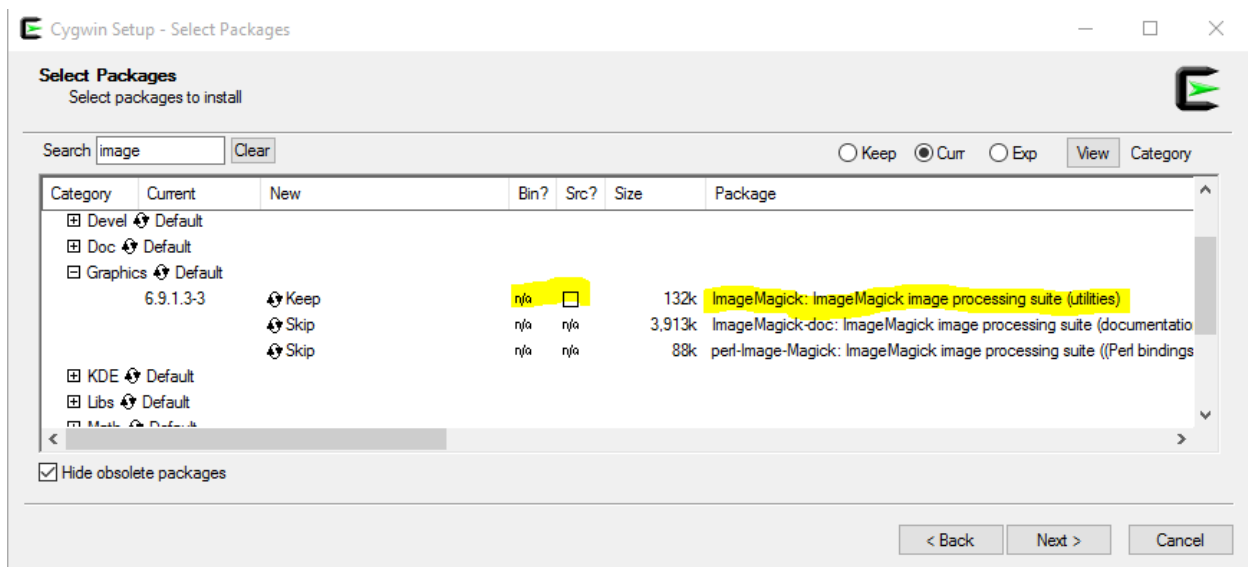
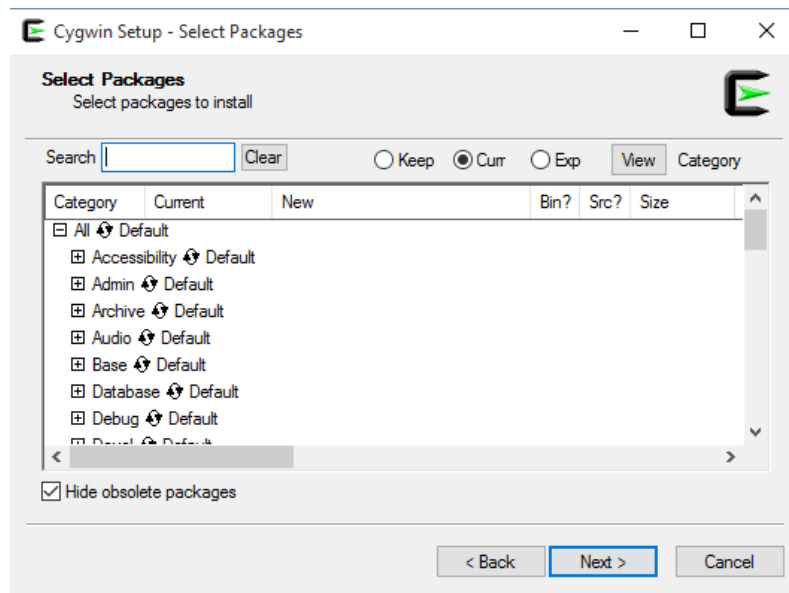


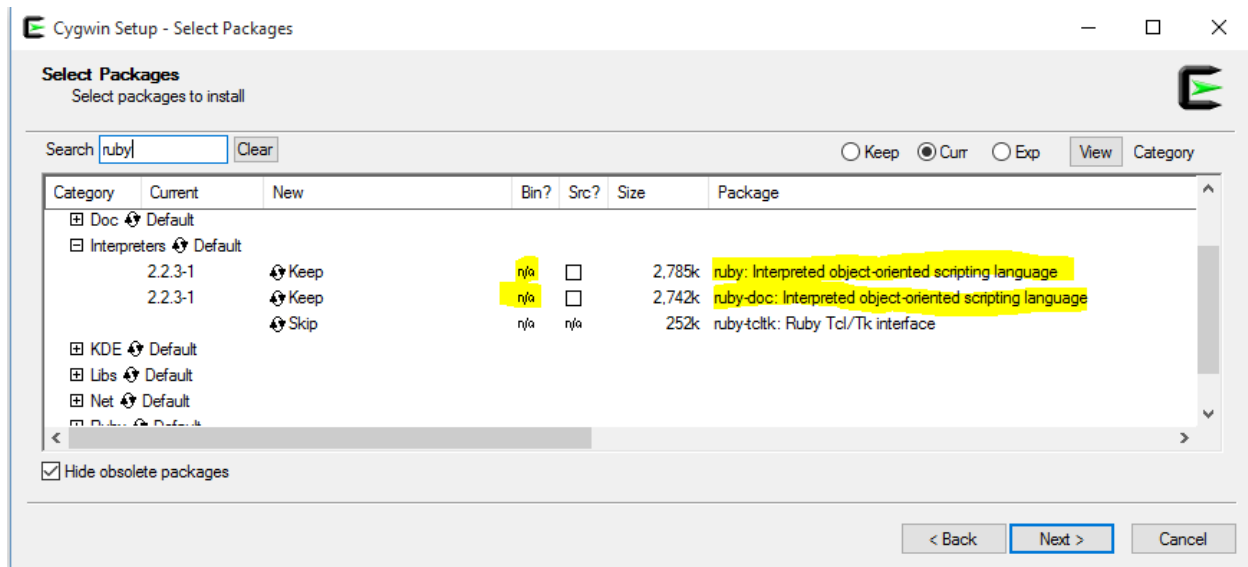


4. Choose any mirror near to you.

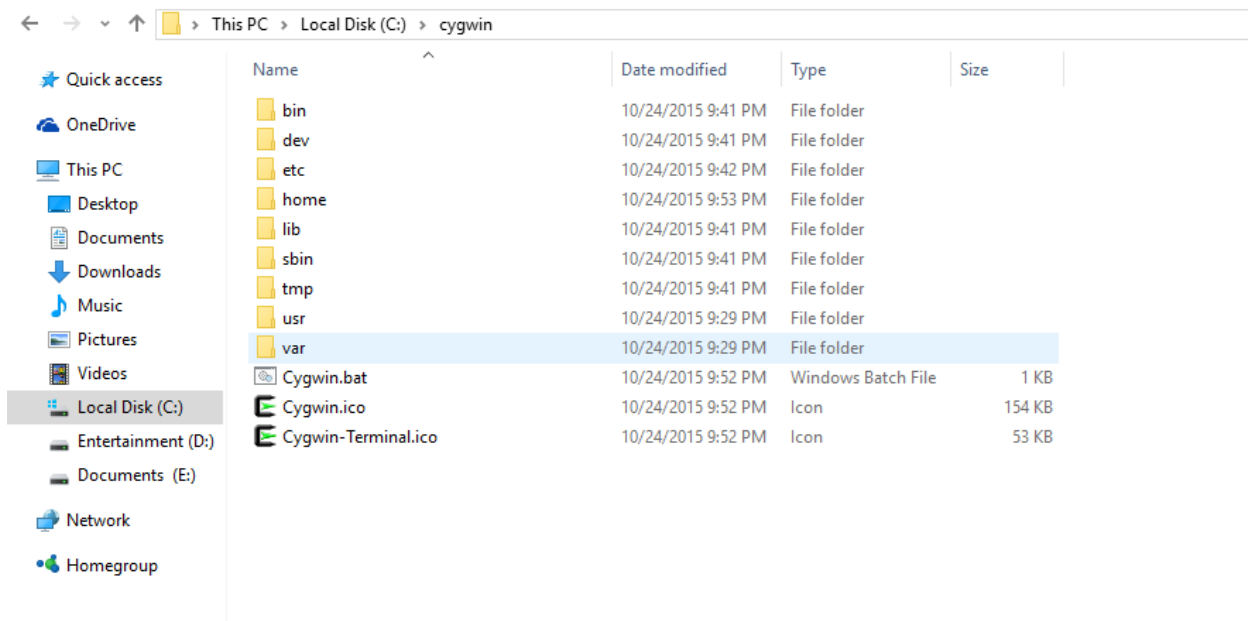


5. Now we need to install two packages. Search for “Image” in the search bar. Then, when the results appear, go to graphics, and then click on ImageMagick. Make sure the “bin” checkbox is checked. “src” is optional.
6. Again search for Ruby in the search bar. Once the search results appear, go to “Interpreters” and select “Ruby” by checking its corresponding “bin” checkbox. These steps are shown below. My steps are different because of the fact that I already have these packages installed.





7. Just click next and wait for the setup to download the required packages and install it.
8. Once Cygwin is successfully installed in your computer, your C:\Cygwin\ folder will look something like this:



9. Run Cygwin.bat once, and then close it once it loads.
10. Now go to home/<username>/
11. Copy your ALL FILES folder here. (The same folder which contains all the DDSM files you downloaded, copies using the command line commands).

**NOTE: MAKE SURE YOUR ALL FILES FOLDER NAME DOES NOT HAVE SPACES IN IT. THIS POSES A PROBLEM SOMETIMES**

Note: Don't worry about the <.1-ddsmraw> files in my case. Those will be generated by our script.

## Setup files and folders for the conversion process

12. Download DDSM software from here - <http://microserf.org.uk/academic/Software.html> Or <https://github.com/trane293/DDSM-Software-Chris-Rose>
13. Clone the repository using Git or download the .zip file from GitHub itself. Extract the contents at a convenient place.
14. Extract my DDSM utility code at a convenient location.
15. Now copy the following files from the newly extracted ddsd-smware folder into my DDSM Utility extracted folder:

**jpeg.exe**

**ddsmraw2pnm.exe**

16. Copy cygwin1.dll file from your C:\cygwin\bin\ folder to my DDSM Utility extracted folder.
17. Your DDSM utility folder will now look exactly like this:

Name	Date modified	Type	Size
pnm	10/27/2015 7:52 PM	File folder	
ConvertDDSMImageToRaw.m	10/27/2015 11:01 ...	M File	5 KB
cygwin1.dll	8/20/2015 3:11 PM	Application extens...	3,399 KB
ddsmraw2pnm.exe	7/18/2009 6:15 PM	Application	485 KB
jpeg.exe	7/13/2009 6:27 PM	Application	116 KB
openDDSMJPEGAndConvertToPNG.m	11/1/2015 6:54 PM	M File	7 KB
openDDSMpngWithOverlay.m	10/27/2015 11:38 ...	M File	3 KB
readBoundary.m	10/24/2015 1:14 PM	M File	4 KB
readme.txt	10/28/2015 12:04 ...	Text Document	2 KB
Tutorial.docx	11/1/2015 8:39 PM	Microsoft Word D...	786 KB
Tutorial.pdf	10/24/2015 1:49 PM	Adobe Acrobat D...	1,860 KB

Note: You will not the .docx version of this tutorial in your folder.

18. Open MATLAB.
19. Open the script “pnmsetup .m” inside the “pnm” folder of your DDSM utility folder.
20. Run the script once. It will complete in a second quietly.
21. Now you are almost ready for the conversion process.

## Open scripts in MATLAB and change paths

22. Open the scripts:

**openDDSMJPEGAndConvertToPNG.m**

**ConvertDDSMImageToRaw.m**

Present in the DDSM utility folder.

23. In the script **openDDSMJPEGAndConvertToPNG.m**, navigate to the first three variables, as shown below:



```

Editor - D:\DDSM Database Downloaded\My Code (v2)\openDDSMJPEGAndConvertToPNG.m
openDDSMJPEGAndConvertToPNG.m
34 % be useful, but WITHOUT ANY WARRANTY; without even the
35 % implied warranty of MERCHANTABILITY or FITNESS FOR A
36 % PARTICULAR PURPOSE. See the GNU General Public License
37 % for more details.
38 %
39 % You should have received a copy of the GNU General
40 % Public License along with this program. If not,
41 % see <http://www.gnu.org/licenses/>.
42 %=====
43 - clear all
44 - clc
45 %% SET THESE PATHS FIRST!
46 - allFilesDirectory = 'D:\DDSM Database Downloaded\DoD Malignant Cases\BCRP 1\PNGFiles\DoD_BCRP_1_ALL_Files\';
47 - writePNGFilesHere = 'D:\DDSM Database Downloaded\DoD Malignant Cases\BCRP 1\PNGFiles\DoDMalignantAllCasesPNG\';
48 - imageOutputFileFormat = '.png'; % Notice the dot. Can be .tif, .jpg...

```

24. Change the first variable to point to the ALL FILES directory where all the DDSM files are copied using the cmd command. Remember you copied this folder into C:\cygwin\home\<username>\ folder. You may, if you want give some other location, but I prefer the Cygwin location.
25. Change the second variable according to your convenience, as to where do you want your PNG files to be written. Please note, the PNG files written are VERY large in size (25-35MB each). The script does not have error checking built in, hence make double sure that you have adequate space where you want to save the PNG files.
26. Change the third variable to write to your own desired format. I highly recommend using the default .png format, however, you may choose to change it, just make sure MATLAB supports it.
27. Now go to the script **ConvertDDSMImageToRaw.m**
28. Now this is the slightly tricky part, and I urge you to pay utmost attention.

```

function imageRAW = ConvertDDSMImageToRaw(filename, columns, rows, digitizer)

%% Prepare and execute command of image decompression
commandDecompression = ['C:\cygwin\bin\bash' -c "cd C://cygwin/home//Anmol//DoD_BCRP_1_ALL_Files//; ./jpeg.exe -d -s ' filename, ''"];
disp('Decompressing LJPEG -> LJPEG1...');
system(commandDecompression);
%// -----
%// Prepare and execute command that convert the decompressed image to pnm format.
rawFileName = [ filename '.1'];
columns = num2str(columns);
rows = num2str(rows);
digitizer = ['' digitizer ''];
commandConversion = ['C:\cygwin\bin\bash' -c "cd C://cygwin/home//Anmol//DoD_BCRP_1_ALL_Files//; ./ddsmraw2pnm.exe ' rawFileName, ' ',
disp('Converting LJPEG1 -> RAW...');
system(commandConversion);
%// -----
%// Write the image into raw format
pnmFileName = [rawFileName '-ddsmraw2pnm.pnm'];
imageRAW = pnmread(strcat(['C:\cygwin\home\Anmol\DoD_BCRP_1_ALL_Files\ ', pnmFileName]));
% imwrite(image,[filename '.png']);
end

```

29. Notice the highlighted region above. Also notice the **double BACKWARD slashes** and **path is in double inverted commas**. In the highlighted region, write your Cygwin bash directory exactly as written above. **If you copy your path from explorer, make sure you put TWO BACKSLASHES IN THE PATH AS SHOWN ABOVE!**

```

function imageRAW = ConvertDDSMImageToRaw(filename, columns, rows, digitizer)

%% Prepare and execute command of image decompression
commandDecompression = ['C:\cygwin\bin\bash' -c "cd C://cygwin/home//Anmol//DoD_BCRP_1_ALL_Files//; ./jpeg.exe -d -s ' filename, ''"];
disp('Decompressing LJPEG -> LJPEG1...');
system(commandDecompression);
%// -----
%// Prepare and execute command that convert the decompressed image to pnm format.
rawFileName = [ filename '.1'];
columns = num2str(columns);
rows = num2str(rows);
digitizer = ['' digitizer ''];
commandConversion = ['C:\cygwin\bin\bash' -c "cd C://cygwin/home//Anmol//DoD_BCRP_1_ALL_Files//; ./ddsmraw2pnm.exe ' rawFileName, ' ', c
disp('Converting LJPEG1 -> RAW...');
system(commandConversion);
%// -----
%// Write the image into raw format
pnmFileName = [rawFileName '-ddsmraw2pnm.pnm'];
imageRAW = pnmread(strcat(['C:\cygwin\home\Anmol\DoD_BCRP_1_ALL_Files\ ', pnmFileName]));
% imwrite(image,[filename '.png']);
end

```

30. Again, notice the highlighted region. Also notice the **double FORWARD slashes**. Here, write your ALL FILES directory path. Again if you copy from explorer, make sure you put **TWO FORWARD SLASHES IN THE PATH AS SHOWN ABOVE**.

```
function imageRAW = ConvertDDSMImageToRaw(filename, columns, rows, digitizer)

%% Prepare and execute command of image decompression
commandDecompression = ['C:\\cygwin\\bin\\bash" -c "cd C://cygwin/home/Anmol/DoD_BCRP_1_ALL_Files//; ./jpeg.exe -d -s ' filename, '''];
disp('Decompressing LJPEG -> LJPEG1...');
system(commandDecompression);
%// -----
%// Prepare and execute command that convert the decompressed image to pnm format.
rawFileName = [ filename '.1'];
columns = num2str(columns);
rows = num2str(rows);
digitizer = [' ' digitizer ' '];
commandConversion = ['C:\\cygwin\\bin\\bash" -c "cd C://cygwin/home/Anmol/DoD_BCRP_1_ALL_Files//; ./ddsmraw2pnm.exe ' rawFileName, ' ', c
disp('Converting LJPEG1 -> RAW...');
system(commandConversion);
%// -----
%// Write the image into raw format
pnmFileName = [rawFileName '-ddsmraw2pnm.pnm'];
imageRAW = pnmread(strcat(['C:\\cygwin\\home\\Anmol\\DoD_BCRP_1_ALL_Files\\', pnmFileName]));
% imwrite(image,[filename '.png']);
end
```

31. Again put here your Cygwin bash path. Take utmost care of the inverted commas and double backslashes.

```
function imageRAW = ConvertDDSMImageToRaw(filename, columns, rows, digitizer)

%% Prepare and execute command of image decompression
commandDecompression = ['C:\\cygwin\\bin\\bash" -c "cd C://cygwin/home/Anmol/DoD_BCRP_1_ALL_Files//; ./jpeg.exe -d -s ' filename, '''];
disp('Decompressing LJPEG -> LJPEG1...');
system(commandDecompression);
%// -----
%// Prepare and execute command that convert the decompressed image to pnm format.
rawFileName = [ filename '.1'];
columns = num2str(columns);
rows = num2str(rows);
digitizer = [' ' digitizer ' '];
commandConversion = ['C:\\cygwin\\bin\\bash" -c "cd C://cygwin/home/Anmol/DoD_BCRP_1_ALL_Files//; ./ddsmraw2pnm.exe ' rawFileName, ' ', c
disp('Converting LJPEG1 -> RAW...');
system(commandConversion);
%// -----
%// Write the image into raw format
pnmFileName = [rawFileName '-ddsmraw2pnm.pnm'];
imageRAW = pnmread(strcat(['C:\\cygwin\\home\\Anmol\\DoD_BCRP_1_ALL_Files\\', pnmFileName]));
% imwrite(image,[filename '.png']);
end
```

32. Again, as two steps above, put here the path to ALL FILES directory present inside your Cygwin/home/<username>/ directory.

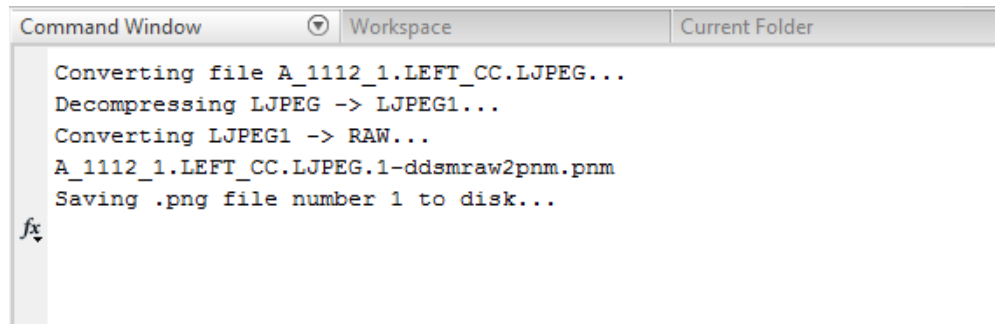
```
function imageRAW = ConvertDDSMImageToRaw(filename, columns, rows, digitizer)

%% Prepare and execute command of image decompression
commandDecompression = ['C:\\cygwin\\bin\\bash" -c "cd C://cygwin/home/Anmol/DoD_BCRP_1_ALL_Files//; ./jpeg.exe -d -s ' filename, '''];
disp('Decompressing LJPEG -> LJPEG1...');
system(commandDecompression);
%// -----
%// Prepare and execute command that convert the decompressed image to pnm format.
rawFileName = [ filename '.1'];
columns = num2str(columns);
rows = num2str(rows);
digitizer = [' ' digitizer ' '];
commandConversion = ['C:\\cygwin\\bin\\bash" -c "cd C://cygwin/home/Anmol/DoD_BCRP_1_ALL_Files//; ./ddsmraw2pnm.exe ' rawFileName, ' ', c
disp('Converting LJPEG1 -> RAW...');
system(commandConversion);
%// -----
%// Write the image into raw format
pnmFileName = [rawFileName '-ddsmraw2pnm.pnm'];
imageRAW = pnmread(strcat(['C:\\cygwin\\home\\Anmol\\DoD_BCRP_1_ALL_Files\\', pnmFileName]));
% imwrite(image,[filename '.png']);
end
```

33. Now here notice that we only a single backward slash and the path is enclosed in single inverted commas. Put here the path to your ALL FILES folder inside the Cygwin/home/<username>/ directory.
34. I suggest you go through this section once again to make sure you did everything right. 99% of the problems that you might face will be due to negligence or errors you made during this section.

## Convert all LJPEG images into LJPEG1 and then into any format like PNG using the supplied MATLAB script.

35. Now the easy part. Run the script **openDDSM LJPEGAndConvertToPNG.m**
36. You will see some text being output to the command window like this:



```
Command Window  Workspace  Current Folder

Converting file A_1112_1.LEFT_CC.LJPEG...
Decompressing LJPEG -> LJPEG1...
Converting LJPEG1 -> RAW...
A_1112_1.LEFT_CC.LJPEG.1-ddsmraw2pnm.pnm
Saving .png file number 1 to disk...
fx
```

37. If you can see this, CONGRATULATIONS! You successfully set up the conversion procedure. Now just sit back and relax. The conversion is a system heavy process, and each image may take from 10 seconds to upto 5 minutes depending on your configuration. On my laptop, each image takes roughly 1 minute to get converted and saved to PNG format on hard disk.