


# SSW URL List Downloading Instructions

## Subsetting Data Sets and Downloading Subsetted Granules

When you Subset Selected Data Sets, once the subset request is successfully completed for a selected data set, the subset results for that data set can be viewed by clicking on the  that appears to the right of the data set description. When all subset requests have been completed, you can view all of the subset results by clicking on the "View subset results" link at the top of the page.

Depending upon which data sets were subsetted, the results may consist of a mixture of links for downloading subsetted files, instructions for obtaining subsets that require offline processing, and explanations in the event that a data set could not be subsetted as requested.

When the results are viewed, each subsetted file can be downloaded by clicking on its link, or they can all be downloaded by using a browser-based download manager. You may also obtain a list of URLs for a data set in Plain Text format by clicking on the "Get list of URLs in a file" link, which will download a file to your local workstation. That file can be used with a Unix command line tool such as `wget` to download all of the subsetted files for that data set. If there is more than one data set, there can be a list of URLs for each data set, so be sure to download the list for each data set.

## Browser-Based Download Manager Usage

Download managers are browser tools which make it easy to download many URLs at once. If you do not already have a download manager installed, you can choose one of these:



How to make the "DownThemAll" Mozilla plugin work effectively for downloads of subsetted data

Users attempting to employ the "DownThemAll" Mozilla plug-in should make the following changes to the default preferences for the plug-in:

- The Multipart download capability should be disabled. Set *Max. number of segments per download* to **1** in the *Multipart download* item, under the *Advanced* tab in the *Preferences* window.
- Concurrent downloads should be disabled. Set *max concurrent downloads* to **1** in the *Download* item, under the *Main* tab of the *Preferences* window.
- Users experiencing network dropouts might consider enabling *Auto Retries* (also under the *Main* tab of the *Preferences* window) by setting *Retry each* to **30 minutes**, and *Max. Retries* to **10**.

Selecting files with a download manager

Download managers can download all URLs that appear on a web page. Although only the URLs for subsetted files will be visible on the subset results page, there will also be some URLs that are not visible, such as the one that points to this page of instructions. To avoid downloading those URLs, make sure that they are not selected in your download manager before starting the download. For example, URLs that end in '.html' point to web pages that you probably do not want to include with your data.

## Unix Command Line Tool Usage: wget

(For use with `wget`, please make sure that the `wget` version is at least 1.18; run `wget -V` to find the version. You can download `wget` [here](#).)

1. If you have not already done so, follow the steps in [How to Register With Earthdata Login](#) and [How to Link GES DISC With Your Account](#).
2. Download the plain text list of URLs from your browser to your local workstation by clicking on the "Get list of URLs in a file" link. The downloaded file will have a name similar to "SSW\_download\_2011-01-01T00:00:00\_12345\_dGoR9Zqq.inp". Note which directory the downloaded file is saved to, and in your Unix shell, set your current working directory to that directory.
3. Create a `~/netrc` file pointing to `urs.earthdata.nasa.gov` and an empty `~/urs_cookies` file by following the instructions for [How To Access Data With cURL And Wget](#).
4. On your command line:

```
wget --load-cookies ~/urs_cookies --save-cookies ~/urs_cookies --auth-no-challenge=on --keep-session-cookies --content-disposition -i
SSW_download_2011-01-01T00:00:00_12345_dGoR9Zqq.inp
```

If step 4 is done without ever having done step 1, the file `~/urs_cookies` will contain incorrect information, in which case you should delete `~/urs_cookies` and create it again (step 3 of [How To Access Data With cURL And Wget](#)).

Sometimes the downloaded files will have names that are the same as the URLs that were used to download them, and the files may even fail to download if that name is too long. Usually the "--content-disposition" option prevents this, but it can happen when the subsetting is using OPeNDAP (in which case the URL usually contains the string "opendap"). A sequence of Unix commands that combine "egrep", "paste", and "awk" can be used to download the files with more reasonable file names, as follows.

Examine the plain text of URLs to determine what the names of the files (the portion following the last slash of each URL) begin with, and what the extension (the portion just before the question mark that begins with a dot) is. For example, let's say that each name begins with '3B', and the extension is '.nc'. Then use these commands to download the files:

```
egrep -o '3B.*\.nc' SSW_download_2011-01-01T00:00:00_12345_dGoR9Zqq.inp > temp1.txt
```

```
paste SSW_download_2011-01-01T00:00:00_12345_dGoR9Zqq.inp temp1.txt > temp2.txt
```

```
awk 'system("wget --load-cookies ~/.urs_cookies --save-cookies ~/.urs_cookies --auth-no-challenge=on --keep-session-cookies --content-disposition \"$1\" -O \"$2\" ")' temp2.txt
```

(In this example, each file started with '3B', and '3B' did not appear anywhere before the last slash. If '3B' did appear before the last slash, we would have needed to make the starting portion longer to distinguish it from what appears before the last slash).

## Unix Command Line Tool Usage Example: curl

The "curl" command is not able to download a plaintext list of URLs without modification. However, if you have perl and wget installed on your Unix host, this script can be used to download a list of URLs:

```
#!/usr/bin/perl

#
# Perl script to invoke wget on a set of OTF-specific URLs contained in one or more URL listing files
# Each URL is parsed to find the LABEL attribute and use that for the local filename
#

use strict;

while ( my $file = shift ) {
    my $n;
    open( IN, $file ) or die "Failure opening $file for reading.\n";
    my @urls = <IN>;
    chomp(@urls);
    close(IN);
    foreach my $url (@urls) {
        $n++;
        my $label;
        if ( $url =~ /HTTP_services.cgi/ ) {
            $url =~ m/LABEL=(.*?)\&/;
            $label = $1;
            unless ($label) {
                $url =~ m/FILENAME=(.*?)\&/;
                my $name = $1;
                $label = ( split /\%2F/, $name )[-1];
            }
        }
        else {
            $label = ( split /\//, $url )[-1];
        }
        unless ($label) {
            $label = "OTF_Download_$file.$n";
            print "No appropriate filename found in URL - using $label\n";
        }
        my $cmd =
            "wget --load-cookies ~/.urs_cookies --save-cookies ~/.urs_cookies --auth-no-challenge=on --keep-session-cookies \"$url\" -O $label";
        ` $cmd `;
    }
}
```

You can use a text editor to create a file downloadUrlList.pl by copying the text of the script and using the paste command of the editor. Note that if perl is installed in a directory other than /usr/bin/perl, you will need to modify the first line of the script accordingly. Save the script in some directory where you have permissions to write a file (e.g. an existing ~/bin directory).

1. If you have not already done so, follow the steps in [How to Register With Earthdata Login](#) and [How to Link GES DISC With Your Account](#).
2. Download the plain text list of URLs from your browser to your local workstation by clicking on the "Get list of URLs in a file" link. The downloaded file will have a name similar to "SSW\_download\_2011-01-01T00:00:00\_12345\_dGoR9Zqq.inp". Note which directory the downloaded file is saved to, and in your Unix shell, set your current working directory to that directory.
3. Create a ~/.netrc file pointing to urs.earthdata.nasa.gov and an empty ~/.urs\_cookies file by following the instructions for [How To Access Data With cURL And Wget](#).
4. If the perl script you created is located in ~/bin/downloadUrlList.pl, then to download the list of URLs contained in the plain text file named SSW\_download\_2011-01-01T00:00:00\_12345\_dGoR9Zqq.inp:

```
perl ~/bin/downloadUrlList.pl SSW_download_2011-01-01T00:00:00_12345_dGoR9Zqq.inp
```

If step 4 is done without ever having done step 1, the file ~/.urs\_cookies will contain incorrect information, in which case you should delete ~/.urs\_cookies and create it again (step 3 of [How To Access Data With cURL And Wget](#)).