

Steps to download the SGDP dataset:

- Get a personal grid x509 certificate to download data using GridFTP from Fermi Lab.
- To get a personal certificate follow the instructions from this link:
https://fermi.service-now.com/kb_view.do?sysparm_article=KB0010815
and use the VO as: SCDA
- Alternatively if you are from an institute included in cilogon (other than google) you can use
 - <https://cilogon.org>
- Once you get your certificate follow the instructions in the email and upload it to your browser, and send the subject (which will look something like /DC=org/DC=cilogon/C=US/O=Google/CN=User Name A16321) and mail it to ifisk@simonsfoundation.org
- Follow instructions from the below link if you will be using Globus tools for submitting grid jobs from Linux/UNIX:
https://fermi.service-now.com/kb_view.do?sysparm_article=KB0010815. Make sure you do this step as soon as you get your certificate and use the same browser window. Note: If you wait too long the certificate is no longer in the PKCS#12 format that you need for this step.
- Install osg-ca-certs and osg-client on your machine; will probably need help from the Systems group to do this. The instructions for this are at: https://twiki.grid.iu.edu/bin/view/Documentation/Release3/InstallOSGClient#6_2_Stopping_and_Disabling_Servi
Note for regular users without root access there is an OSG tarball option:
<https://twiki.grid.iu.edu/bin/view/Documentation/Release3/InstallOSGClientTarball>
- Send the certificate to Yujun Wu (yujun@fnal.gov) or Dmitry O Litvintsev (litvinse@fnal.gov) to ensure that things are set up properly.
- Run the following command: `./opt/globus-5.2.5/etc/globus-user-env.sh`, to ensure you are running the correct version of globus
- Run the command `grid-proxy-init -valid 168:0` (This will allow keep the proxy active for a week; after which you will need to renew it again)
- Test if the download is working using the following command:
`"globus-url-copy -vb -dbg -nodcau gsiftp://fndca1.fnal.gov:2811/temp/testfnal.txt file:///tmp/testfile"`
- Copy a file called COMPLETE_FILE_LISTING in your folder using the below command:
`globus-url-copy gsiftp://fndca1.fnal.gov/COMPLETE_FILE_LISTING file:///^pwd`/COMPLETE_FILE_LISTING`

- Copy the script `complete.sh` (see below) in the same folder as where you have the `COMPLETE_FILE_LISTING` file; and run `./complete.sh` to copy all the files.
- Script: `complete.sh`:
`#!/bin/bash`

```
cat COMPLETE_FILE_LISTING | grep SGDP | while read path size cksum
do
# echo "globus-url-copy -c gsiftp://fndca1.fnal.gov${path} file:///^pwd`/${path}"
  globus-url-copy -c -vb -nodcau -cd -bs 2000000 -sync gsiftp://fndca1.fnal.gov${path}
  file:///^pwd`/${path}
done
```

- Once you have succeeded, you will find the transfers are much faster if parallel streams are enabled. You may need to speak with the local network administrator to open ports in the firewall

Two environment variables need to be set

```
export GLOBUS_TCP_PORT_RANGE=50000,50100
export GLOBUS_HOSTNAME=Name_of_the_external_IP
```

And ports 50000-50100 need to be open in the firewall

Then add “-p 10” to the list of options in the `globus-url-copy` command above