

GEO submission

version April 01, 2021

1. Gather files in a single directory

- raw files for each sample
- individual or combined final file(s)
- manuscript associated with data (if available) or put together some basic information like title, authors, methods

2. Run checksum

For all raw and final data files, you need to provide an md5 checksum value to ensure data integrity. Navigate to your directory with the raw and processed files and run

```
# LINUX
md5 * > checksum.txt
```

```
# MAC
md5sum * > checksum.txt
```

3. Fill out metadata

Find your data type at <https://www.ncbi.nlm.nih.gov/geo/info/spreadsheet.html#seq> and download the template metadata file. For example, RNA-seq is [here](#). Fill out all required fields using the information you gathered above.

4. Create upload space

Log-in to your NCBI account. Go to <https://www.ncbi.nlm.nih.gov/geo/info/submissionftp.html> and create an upload space (Step 1 on the page). Refresh the page until a path appears.

5. Upload to GEO

From AWS S3 If your data files are already on S3, it is much faster to upload them directly from there. Use an EC2 instance for the best speed. I think even the free EC2 tier would work.

Make sure all the raw, final, and metadata files are in a single directory with nothing else. Then log-in to your instance and sync (not copy) the bucket to an empty directory. For example

```
mkdir geo_submission
sudo chmod 777 -R geo_submission
```

```
s3fs [BUCKET_NAME_HERE] geo_submission -o passwd_file=~/.passwd-s3fs \
-o default_acl=public-read -o uid=1000 -o gid=1000 -o umask=0007
```

```
cd geo_submission
```

```
#cd further into subdirectories if your files are sequestered in one
```

Upload files to GEO (Step 2 on the page)

```
sftp geoftp@sftp-private.ncbi.nlm.nih.gov
password: [PASSWORD HERE]
cd [PATH like uploads/kdillmcfarland_qXMN4a4g]
mkdir new_geo_submission
cd new_geo_submission
put *
```

You can name the directory something other than `new_geo_submission`. In fact, if you're uploading multiple projects, you'll have to put them in separate directories here.

From local computer

I have not done this so please feel free to add this section if you do!

6. Notify GEO

Once all your files are uploaded in the terminal, click "Notify GEO" and fill out the form (Step 3 on the page). Submit! You will receive a confirmation email right away. Submissions are reviewed within 5 business days.