# Definitions:

BioProject – The overall study

BioSample – An experiment within the study (can consist of multiple samples)

SRA submission – A sample within an experiment

Run – Sequencing reads for a sample

# Setting up a submission:

1. Create a BioProject at [https://submit.ncbi.nlm.nih.gov/] (it will take a day or two for the project to get assigned an accession number).
2. Create one or more BioSamples within the BioProject (again, may take a day or two).
3. Create new SRA submissions at [http://www.ncbi.nlm.nih.gov/Traces/sra\_sub/] and add runs.
   1. Batch submissions are also possible [https://submit.ncbi.nlm.nih.gov/subs/sra/].
4. [optional] Check md5sum of files to be uploaded by navigating to the directory on MSI and using the command [md5sum filename1] or [md5sum \*.\*] for the entire directory.

# Submitting sequence read files (individual samples):

1. Access the NCBI FTP site using the command (in terminal on MSI)

$ lftp -u subftp,password ftp-private.ncbi.nlm.nih.gov

1. Navigate to your directory as described in the SRA submission instructions and create a directory for your submission. Enter this directory,
2. Navigate to the directory with your sequencing files using [lcd /path/to/sequences].
3. Upload files using command [mput \*.\*] to upload all sequences in the directory or [mput K\*], for example, to upload all samples starting with K.
4. If the NCBI connection times out, reestablish it using command [open ftp-private.ncbi.nih.gov].

# Submitting sequence read files (batch submission):

1. Access the NCBI FTP site using command [ftp ftp-private.ncbi.nlm.nih.gov].
2. Login in to the sequence read archive using the name [subftp] and enter password provided on the run files webpage.
3. Navigate to your folder [cd uploads/yourfolder]. The path is provided by NCBI.
4. Make a new folder for uploads [mkdir folder1].
5. Navigate to this folder [cd ./folder1]
6. Navigate to the directory with your sequencing files using [lcd /path/to/sequences].
7. Turn off interactive mode using the command [prompt].
8. Upload files using command [mput \*.\*] to upload all sequences in the directory.
9. If the NCBI connection times out, reestablish it using command [open ftp-m].
10. Fill out the batch submission form (using a different form for each BioProject).