Instructions on how to download Rockland Sample imaging data from AWS

There are two options on how to download neuroimaging data from the Rockland Sample. Option 1 is based on a *bash* script and option 2 is based on *python.* While option 1 is a simpler to use (does not depend on python), option 2 is useful if you would like to download a large amount of data. We recommend reading through both methods before selecting which option is best for you.

Option 1: Download using a bash script:

1. You will need to install the lasted version of AWS's Command Line Interface
   You can follow step-by-step instructions on AWS's website here.

   Note: for using our download script you do not need to perform a setup, such as setting up the AWS Access Key ID or AWS Secret Access Key.

2. Next, you will need to download our "NKIRS_download_links.sh" script. You can download it here.

3. Download the full list of links to the files on AWS. You can download the "aws_links.csv" file here.

   An alternative would be to download a list of AWS links using our LORIS database. In the LORIS database, you can generate a list of AWS links on the "MRI" instrument.

4. If you downloaded aws_list.csv, you would need to modify the file to only have the rows of files that you would like to download. You will also need to remove the header of the CSV file and any column that is not an s3 link. For example, you will need a file that looks like this:

5. Now you are ready to download the files. You will need to open a terminal and go to the location where you saved the "download_links.sh" script. Next, run the following command:

bash NKIRS_download_links.sh -i list_with_S3_links.csv -o ./DOWNLOADFOLDER/

Note, make sure you create "DOWNLOADFOLDER" before running the script and that the list with the "list_with_S3_links.csv" is in the same location of the script or you include the full path.

While downloading the files, you can ignore the warning:
" Unverified HTTPS request is being made to host 'fcp-indi.s3.amazonaws.com'. Adding certificate verification is strongly advised. See: https://urllib3.readthedocs.io/en/latest/advanced-usage.html#ssl-warnings"
Due to some problems with firewalls, we added a flag on the script to not verify SSL certificates when downloading data. If you do not have a firewall problem when downloading data from AWS and want to remove this warning, you can modify the download script. You will need to go to the line where it is written:

aws s3 cp ${filepath} ${output}/${exactPath} --no-sign-request --no-verify-ssl

and remove "--no-verify-ssl". After removing the "--no-verify-ssl" flag, you should not be receiving this warning anymore.

Option 2: Download using python script:

1. The Python script located here allows you to specify a number of options so to download a customized subset of the sample from the S3 bucket. In particular, it allows you to specify:

   - Age Ranges
   - Sex
   - Handedness
   - Study and Visit
   - Scan Type
   - Series

   If none of these items are specified, the script will assume that you would like to download all available raw NKI-RS Neuroimaging Data release data without derivatives.

   The script requires the following dependencies to function properly:

   - Python version 3.6
   - pandas
   - boto3

To install pandas and boto3 within Python, you can use the following command: *pip install pandas boto3*

2. If you want to specify which studies / visits to download, you must take a study code and append the desired visit code to it, and then add this after the *-v* option. You may specify multiple *study/visit code combinations* (using the NITRC/AWS convention) after this option.

For example:

```
python download_rockland_raw_bids_ver2.py -o /data/output -v BAS1
```

This will download all of the files for the Baseline visit and store them in /data/output in BIDS format.

3. Multiple scan types, separated by spaces after the scan type parameter, can be selected. These scan types can take on the following values, which come from the BIDS specification:

- anat - Anatomical scans.
- func - Functional/resting-state scans.
- dwi - DTI scans.

For example:

```
python download_rockland_raw_bids_ver2.py -o /data/output -t anat dwi
```

This will download all of the anatomical and dti data only across all visits.

4. Similarly, multiple series can be specified at the command line. These series can take on the following values:

- Non-neurofeedback studies:

  - REST645
  - REST1400
  - RESTCAP
  - RESTPCASL
  - CHECKERBOARD645
  - BREATHHOLD1400
  - CHECKERBOARD1400

For example:

```
python download_rockland_raw_bids_ver2.py -o /data/output -e REST645
CHECKERBOARD645
```

This will download only the data for these two series across all participants and visits.

5. Here is a more complex example of the downloader script's syntax, which illustrates how you can combine command line options to generate extremely nuanced subsets of the Rockland sample.

```
python download_rockland_raw_bids_ver2.py -o /data/output -e REST645 -v
BAS1 -t anat func -gt 15 -lt 40 -m L -x F
```

This command would download all anatomical series and all resting state series with a 645 ms TR for left-handed women between the ages of 15 and 40, but only from the abridged and second visits of the discovery science and neurofeedback studies. This data could be used to assess within-subject variations in connectivity and neuroanatomy for this specific population, since the first visit for the discovery science and neurofeedback protocols are identical.

To get more information on all of the parameters you may use, type:

```
python download_rockland_raw_bids_ver2.py -h
```