CONSORTIUM FOR
AMC Advancing Management and Prevention of Cancer in People with HIV
DATA COMMONS

Data Upload – Linux OS

Data upload to the AMC Data Commons utilizes Amazon Web Services (AWS) S3 storage. Below are instructions for uploading data files to the AMC Data Commons AWS S3 storage via two available Linux methods:

- FileZilla Pro
- AWS Command Line Interface

For instructions on data uploads from Windows or Mac systems, please reference the other system specific instructions documents. If you have any questions or require assistance, email amcdc_data@emmes.com for support.

Before uploading, ensure data is appropriately encrypted and compressed (See AMC-DC Data Preparation Instructions for details).

FileZilla Pro

1. Download FileZilla Pro for Linux. The program is available for download at the following link:

   FileZilla Pro for Linux

   To receive reimbursement from AMC for the software, submit an invoice to Emmes at invoices@emmes.com. One license per institution is eligible for reimbursement.

2. In the menu bar, click on File > Site Manager….

3. Click on New Site.

4. Select S3 – Amazon Simple Storage Service from the Protocol drop-down list. FileZilla Pro will automatically fill in the Host field (s3.amazonaws.com). Leave port as-is.

5. Copy the Access Key ID and Secret Access Key into the corresponding fields in Site Manager.

   AWS Access Key: *** (This will be provided separately for security purposes)
   AWS Secret Access Key: *** (This will be provided separately for security purposes)

6. Click on Connect. AWS file storage buckets will be displayed in the Remote directories section of the FileZilla Pro window (right panel). Local directories are displayed in the left

panel. To transfer files, navigate in the left panel to your local directory containing the files to be transferred, then click and drag files over to the right panel.

Note: These instructions are available in video form at https://youtu.be/Ea6WVWADQJc

AWS Command Line Interface

1. Install and configure the Amazon AWS Command Line Interface (CLI) locally on your machine. This requires Python version 2.6.5+, or Python version 3.3+. Detailed installation instructions can be found at:

   http://docs.aws.amazon.com/cli/latest/userguide/installing.html

   If you have a supported version of Python and the Python package manager (pip) installed, you can install the CLI with a single command:

   $ **pip install awscli --upgrade --user**

   If you don't have Python and pip, the installation instruction link above provides an operating system specific guide at the bottom of the page for obtaining them.

2. After installing the CLI, run the following command to verify the installation was successful:

   $ **aws --version**

   If the machine is unable to find the program after installation, add the path to the executable file to your PATH variable. Reference the installation instruction link above for operating system-specific guides to complete this update.

3. Next, configure the CLI to be able to upload to the AMC-DC S3 bucket. In the AWS S3 system, the term "bucket" refers to a storage folder. Run the following command:

   $ **aws configure**

   Enter the following information:
   AWS Access Key: *** (This will be provided separately for security purposes)
   AWS Secret Access Key: *** (This will be provided separately for security purposes)
   Default region name: us-east-1
   Default output format: json

4. To upload a file, use the following command, replacing **[filename]** with the file you are uploading and **[project-specific URL]** with the bucket name provided by ODMC staff:

   $ **aws s3 cp [filename] [project-specific URL]**

   Wild cards may be used as part of the upload file name to upload multiple matching files:

   $ **aws s3 cp *[end of file name] [project-specific URL]**