



Data Preparation – Windows OS

Data uploads to the AMC Data Commons utilize Amazon Web Services (AWS) S3 storage. Below are instructions for compressing and encrypting data files prior to uploading to the AMC Data Commons AWS S3 storage.

For instructions on Mac and Linux 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 completing data preparation per these instructions, ensure:

- A) data submitted are free of PHI/PII**
 - a. for images, please ensure files and file headers only use Participant ID or Specimen ID, per the Imaging – Pathology Data Standardization Plan.**
 - b. for sequencing, use GlobalTrace specimen IDs to identify samples and ensure sequencing files are encrypted prior to upload per instructions below.**
- B) an additional backup of all original files is maintained.**

The following free and open source software will be required to prepare data files:

[7zip download instructions](#)

[OpenSSL download instructions](#)

1. File Preparation

Confirm that the files to be transferred are saved in the appropriate format. CSV files should be saved as standard Comma Delimited files (not CSV UTF-8). To save an Excel spreadsheet as a CSV file, navigate to Save As, and revise the File Type from Excel Workbook (.xlsx) to CSV (Comma delimited) (*.csv).

2. SHA256 Sums

Calculate SHA256 sums for each original uncompressed file to facilitate confirmation of file integrity after the file has been uploaded to the S3 bucket, decrypted, and decompressed. Enter the following into the command prompt to calculate the SHA256 sum for a single file:

CertUtil -hashfile <path to file> SHA256

Save the generated SHA256 sum for each file into a single .txt file and include with the file collection to be uploaded.

3. Compression

Compress the files into a gzip format using 7-Zip:

- a. Open 7Zip File Manager.
- b. Click on the file you wish to compress and click “**Add**”. This will identify the file that will be compressed.
- c. “**Add to Archive**” window will appear. Ensure that the Archive format is set to “**gzip**” and then select “**OK**”.
- d. 7Zip will now process the file and compress it in a gzip file that’s located in the same destination where the original file is.

4. Encryption

Note: The following commands should be executed within the Command Prompt, which can be accessed by searching ‘Command Prompt’ from the Windows Start menu.

Encrypt the compressed files using openssl. A password that will be used for encryption & decryption will be needed. ODMC staff will share the encryption password separately. The following command encrypts a single file:

```
openssl enc -aes-256-cbc -md sha256 -pass pass:<password> -in <filename>.gz  
-out <filename>.gz.enc
```

Please use the .enc file extension postfix for all encrypted files. Please add the .enc extension to any existing extension, e.g. for zipped files, please use <filename>.gz.enc