



AGI Data Download & Delivery Hosted @ The University of Arizona

Table of Contents

Introduction.....	2
⚠️ DATA RETENTION NOTICE ⚠️	2
Downloading with FileZilla.....	3
Downloading with Wget.....	4
Data Verification.....	5
Other Delivery Methods.....	6
Cloud Storage.....	6
Globus.....	6
Cyverse.....	6
Mail.....	6
Sending data to AGI.....	7
Appendix.....	8
AGI FTP Certificate.....	8

Introduction

AGI typically delivers clients their data via our secure FTP service, connected to the [Internet2](#) backbone at 10-gigabits/second. [File Transfer Protocol](#), although an old protocol (around since 1971), is still robust enough for today's needs and supports encryption. Unfortunately, the implementation in web browsers is lacking many of the features needed to download data securely, quickly, and efficiently.

Therefore, we recommend that you use an FTP program, not a web browser.

Here we provide some suggestions and instructions for downloading your data. The instructions can be adapted for other FTP programs or downloaders you may wish to use instead.

Besides FTP, we also have [other delivery methods](#) available.

⚠ DATA RETENTION NOTICE ⚠

Out with the old , In with the new

Please keep the following in mind. Beginning on the date the Administrator notifies you your data is ready for download (**T**), it will be subject to removal from our systems according to the following schedule:

- **T+1** month: all data purged from FTP storage.
- **T+2 months = S**: subread data purged from primary storage.
- **S+14 days**: subread data purged from backup storage.
Subread data completely gone/unrecoverable at this point.
- **T+6 months = H**: HiFi/CCS & other analysis data purged from primary storage.
- **H+14 days**: HiFi/CCS & other analysis data purged from backup storage.
HiFi/CCS & other analysis data completely gone/unrecoverable at this point.

Therefore, **if you need more time to retrieve and/or validate your data, please notify us as soon as possible.** We are happy to extend further time to you. You may reply to the e-mail or contact the administrator directly at: ftpadmin@genome.arizona.edu

Downloading with FileZilla

FileZilla is a free, cross-platform FTP client. It supports secure FTP (FTPS), simultaneous downloads, and automatic resume. It can be downloaded for Windows, Mac, and Linux from <https://filezilla-project.org/> (FileZilla Client). After installing and launching the app:

First, click Edit > Settings..., then Transfers. Set Maximum Simultaneous Transfers to 10 (or lower if you are on a slower computer or network connection), and both the Limit for concurrent downloads and Limit for concurrent uploads to 0. Uncheck Enable speed limits. Now go to the next section, FTP: File Types, and change the Default transfer type to Binary. Click OK.

Second, click on the Site Manager button in the upper left.

Use the following settings:

- Protocol: FTP
- Host: ftp.genome.arizona.edu
- Port: 21
- Encryption: Choose one of the following options:
 - Require explicit FTP over TLS ✓ (recommended)
This option will use encryption for login and transfer of your data, eavesdroppers won't be able to easily obtain your credentials nor data.
 - Only use plain FTP (insecure) ⚠ (not recommended)
This option will **not** use encryption. Your transfer speed may be faster but your login credentials and data are easily obtained by eavesdroppers.
- Logon type: Either Normal or Ask for password. Normal allows you to save the password, while Ask for password will always ask.
- User and Password: provided in the e-mail.

Third, click on the Transfer Settings tab, uncheck Limit number of simultaneous connections.

Finally, you may click the Connect button. If you chose to use encryption, you'll be presented with AGI's FTP certificate for verification. To ensure an eavesdropper is not tricking you to send them your login credentials or data, it's important to verify the fingerprints match:

```
SHA1 Fingerprint = CD:1C:7F:54:36:F7:5A:A5:F5:71:DC:2E:DF:21:66:C9:C1:21:77:16
SHA256 Fingerprint = E0:7F:4D:4E:FC:09:EF:53:F1:00:94:BB:A9:29:B3:50:A5:C7:B1:54:2D:7D:D9:5D:F8:
0B:A2:A3:32:DD:F8:7A
```

Once verified, the certificate can be remembered for future connections.

Now that you are connected, you can select and drag & drop the files or folders you wish to download from the right half of the window (which shows the available data) over to the left half of the window (which shows your local file system). The bottom portion shows the transfer queue.

★ **If the connection is getting hung up, try going back into the Site Manager > Transfer Settings tab and choose Active for the Transfer Mode.**



Downloading with Wget

Sometimes customers need to download their data to a computational server for further analysis, and these servers only support a command line interface. Therefore, we provide instructions for using Wget, also a free, cross-platform downloader that supports secure FTP (FTPS) and automatic resume.

Wget is typically already installed on Linux servers. If it isn't, you can typically install it from your package manager, or ask the administrator to install it. If neither of those is an option, you can download the source code from <https://www.gnu.org/software/wget/> and compile it manually.

Example Wget command with encryption ✓ (recommended)

```
wget -c -m --secure-protocol=TLSv1_2 \  
--ca-certificate=/path/to/ftp_genome_arizona_edu_cert.cer \  
--user=<Username from e-mail> --ask-password \  
ftps://ftp.genome.arizona.edu/*
```

Use the [AGI FTP Certificate](#) for the `--ca-certificate` option.

This command will: (c)ontinue getting a partially-downloaded file, (m)irror (download everything in) your FTP account, and use encryption.

Older versions of Wget do not support FTPS/encryption. In this case, it is recommended to update to the latest version, or ask the administrator to update. Without encryption, your login credentials and data are easily obtained by eavesdroppers.

Example Wget command without encryption ⚠ (not recommended)

```
wget -c -m --user=<Username from e-mail> --ask-password \  
ftp://ftp.genome.arizona.edu/*
```

★ **If the connection is getting hung up, you may need to add the `--no-passive-ftp` option (for example, after the `--ask-password` option).**

❗ **WGET, with the `-m` option, will retain a hidden `.listing` file in every directory. They are primarily used for verifying all data was downloaded, and are normally removed after a successful download. Since they are generated by WGET and not part of our data delivery, they can interfere with [data verification](#) and will show up as errors.**

If you'd rather not keep these files, replace the `-m` option with: `-r -N -l inf`

Data Verification

Once you've finished downloading your data, it's important to verify you actually received the data we're providing to you. Due to the unstable nature of the Internet, it is common that the data download can be interrupted. This can cause an incomplete download or corrupted file.

To verify this hasn't happened, we provide checksums for all of the data in the `sha1sums.txt` file(s). To process `sha1sums.txt`, use a program like `sha1deep` (recommended) or `sha1sum`. `sha1deep` is part of the `hashdeep` application (Linux, Mac) and uses multiple threads for verification. It can be installed from your package manager or downloaded from <https://github.com/jessek/hashdeep>. `sha1sum` is typically installed by default on Linux and Mac, but it only uses one thread so it takes quite a bit longer.

For Windows, you may need to search for a SHA-1 checksum program.

Note: to verify large files such as `subread bam` and `scraps data`, it can take hours for the verification to complete, depending on your computer's processing power. Please allow the verification to complete so you know for sure the integrity your data.

Example `sha1deep` command:

```
sha1deep -er -wx sha1sums.txt *
```

This will show a (convoluted) progress indicator as it reads through your data, but you can get an idea that it is running and how much time is left. It is normal to see a `sha1sums.txt` does NOT match output. If there is no further output, then it indicates the data verification was successful and your data download is complete. If you see any other does NOT match output, then it indicates an incomplete download or corrupted file. You'll want to try resuming the data download or deleting the affected file and starting over. However, if you deliberately did not download all of the provided data, then you'll see a corresponding does NOT match output for each skipped file.

Example `sha1sum` command:

```
sha1sum -c sha1sums.txt
```

In this case, there is no progress indicator (you can check your process manager or system monitor to verify it is still running). If it completes without any FAILED output, then it indicates the data verification was successful and your data download is complete. If you see any FAILED output, then it indicates an incomplete download or corrupted file. You'll want to try resuming the data download or deleting the affected file and starting over. However, if you deliberately did not download all of the provided data, then you'll see a corresponding FAILED output for each skipped file.

i If you downloaded using our `WGET` command, the `-m` option causes a hidden `.listing` file in every directory to be retained. Since they are generated by `WGET` and not part of our data delivery, they can interfere with data verification and will show up as errors. These can be ignored, and/or the `.listing` files deleted.



Other Delivery Methods

Besides FTP, we can deliver your data using a number of different methods, briefly described below. In any case, please contact us for further details and coordination.

Delivery Method	Data Transfer Options
Cloud Storage	<ul style="list-style-type: none"> AGI Upload to Microsoft Azure AGI Upload to Amazon S3 Possibly: AGI Upload to other cloud storage services
Globus	<ul style="list-style-type: none"> You Download from our Globus endpoint AGI Upload to your Globus endpoint
Cyverse	<ul style="list-style-type: none"> You Download from our Cyverse storage AGI Upload to your Cyverse storage
Mail	<ul style="list-style-type: none"> AGI Fedex to you on a physical disk

Cloud Storage

We have successfully uploaded data to Microsoft Azure and Amazon S3. We could probably upload to other services as well. This can go very fast, as the transfer is done in chunks and will use all available bandwidth.

You will have to setup a storage area that we are allowed to access from our data storage server.



You can download your data from our Globus endpoint.

Alternatively, we can upload your data to your Globus endpoint. You'll need to grant us access to an area on your endpoint. We can upload to your corporate/institutional Globus Connect Server endpoint, or to your Globus Connect Personal endpoint on your desktop or laptop.



You can download your data from our Cyverse storage. You can use Cyberduck (Windows, Mac), iRods/iCommands (Linux), web download, or copy it to your own Cyverse storage area.

Alternatively, we can upload your data to your Cyverse storage. You'll need to allow us write access to a directory on your storage.



We can Fedex your data to you on a physical disk using next-day or overnight service.

For this option, there would be an additional fee for the: physical disk, extra labor involved, mailing of the disk to you, and administrative overhead. This fee can be reduced if you have your own disk you would like us to use instead, just mail it to us first. We can optionally format the disk to a number of different file systems and also use encryption (if the file system or disk supports it).

Sending data to AGI

To send data to AGI, the recommended method is to [use FileZilla](#): just drag & drop files and/or folders over to our directory (may need to request permission from AGI first). Please include checksums for the data you are sending.

We can also receive data via: Globus, Cyverse, or Mail. Please contact us for further details.

If you have to use a command line interface on Linux, we provide one example here using LFTP. Like Wget, LFTP can be installed from your package manager or by an administrator, or directly from <https://github.com/lavv17/lftp>.

To start using LFTP, first create a configuration file in your home directory, `~/.config/lftp/rc`, with the following entries:

```
set ftp:ssl-allow true
set ftp:ssl-auth TLS
set ssl:priority +TLS1.2
set ssl:ca-file /path/to/ftp_genome_arizona_edu_cert.cer
```

Use the [AGI FTP Certificate](#) for the `ssl:ca-file` option.

Now `cd` to the directory where the files or directories you want to send are contained, and log in to our server with the command:

```
lftp -u <Username from e-mail> ftp.genome.arizona.edu
```

You will be prompted for the password. If all goes well, then you should now see a prompt like:

```
lftp user@ftp.genome.arizona.edu:~>
```

To upload individual files, use the `put` command as follows:

```
put -c <lfile>
```

where `<lfile>` is the name of the local file you want to send, with option: (c)ontinue failed uploads.

To upload an entire directory, use the `mput` command as follows:

```
mput -c -d -P 10 <directory>/*
```

where `<directory>` is the name of the local directory you want to send, with options: (c)ontinue failed uploads, create matching (d)irectory structure, and send (10) files in (P)arallel.

Again, please be sure to include the checksums for the data you are sending.

Appendix



AGI FTP Certificate

For Wget download or LFTP upload, the certificate required in order to use encryption has been attached to this PDF. Look for "Attachments" or a paperclip icon and you should see ftp_genome_arizona_edu_cert.cer. Save this file to your Wget or LFTP server and specify it as the argument to the command's certificate option.

Alternatively, use the data below and copy/paste it into a new plain text file named ftp_genome_arizona_edu_cert.cer. Include the BEGIN and END lines with all of the dashes, making sure not to add anything else, such as spaces or new lines.

```
-----BEGIN CERTIFICATE-----
MIH2TCBkGgAwIBAgITQ506u1PXF71K8rLSLjkFeDANBkqhk1G9w0BAQWfADBE
M0swC0YDV0QGEwJVU2ESMBAQA1UEChMjSW50ZXJ1ZDQvMSEwHwYDVQDExnJbknV
bW1vbnB3SU0EGDU2vdmVvIEBID1wheCmH1QwMzE1MDAwMDAwWhcNMTUwMzE1MjM1
OTU5WlBkM0swC0YDV0QGEwJVU2EQMA4GA1UECmBHOXJpem9uYU1EIECAQA1UEChMz
VGH1IFVuaXZ1cnNpdHkgb2YgQXJpem9uYU1EIEB0GA1UEAxMwZm9uZHM1b290ZS5h
cm16b25hLmVkdDCCAIUW0YJKozIhvcNAQEBBQADgEPADCCAgQcggIBANZ0zhv
oX9WDx0BQhBmp1K4awEtnwsePNe0jwT0mQ+72wAbvM0g8Fv6bSt7J8gpAfCrvd
7E11h/QkpvmxW2sH8D30wsWdLUu0Bn91hXtDFksFegeDuuSYTlUP77CID1zoYBv
x2Juxr1LJjxY9DM1q+veNBGq14aSec0oS0LmfE3G9P67zWwHb7a/Ix11hL5B/0
v/JDbc9ZHJc6dp9nTYAKcBac89sYuv33wb+bbms04HK1N1K160MSEHfUrkp
1FK1jz4Cq08VhhJUZecGxgm9ZsIE0Ztqo5I55BPaseGtGUDs8QfG0w0t123FP5x
p9SW6RmJ7SR0weh4GW4AE0n1k5vK4FAVS4t8TA1jX1WdVcrDMZ4rSxnmWdWbCRH
An9NwckTRu5Rt1vyaOpLZ7BVkAgGxc1JlwaNzck7aRF504vpuvJwT5b8wLHQ/rjz
pDvYv8d7fwc1VR8T5H450mDSFU7RGdp5m4JpYwTndB9Je4W7gV+gs0hVUyLcdcz
CKF10EEDJMaXN1duUvLe01Jz1CVSfjv1qCDGUK71iUD0UUAJLpVpBe1dBMppTGGZ1
FD60DFeYZxVnX1f0A9eZZeDmfw+eXOVLfk6/9ZdE1rASXmIteZ5eN6SgA1aGUE
1j62AcCdCgFW5LP0uXGfc68SIZQxNa0NeJwvAgMBAAGIggMIIDITAFB@NVHSM
GDWAgBTvTACSpvtL2L16V4s1FhxS21U312TAgBNVHQ4EF0UNO/0/HXKTSK1Yt7
uGVEw4JOB5UwDeYDVR0PAQH/BAQDAgWcMAWGA1UdEwEB/0CMAAwHQYDVDR01BBYw
FAV1KwYBBQUHwEGCCsGAQUFBwMCMERGA1UdIARCMCAwNAYLKwYBBAGVQECAMCw
JTA1BgggrBgEFBQcCARYXaHR0cHM6Ly93ZWNoaWdvLnVybS9DUFMwCAQZ4EMAOIC
MEAGADQhW05DcnNaAZODGGL2h0dHA6Ly91cmwuc2V1dGlnbn5jb20vSW50b2Z1t
b25u0fTZXZ2ZXJ0Q1IuY3JmHAGCCSQAQUFBwEBAgQYAA7BgggrBgEFBQcCwAoYv
aHR0cDovL2NvY25hZWNoaWdvLnVybS9DUFMwCAQZ4EMAOICMEAGADQhW05DcnNa
1wY1KwYBBQUHwAGGF2h0dHA6Ly9vY3NwLnN1Y3RzZ28uY29tMCEGA1UdEQQoMBIc
FmZ0cC5nZW5vbnV1eXJpem9uY5S1ZHUwgf+BgorBgEEAgZ5AgQCBIIBbgCAwBo
aAB2AM8RvU7VLnvy84db2Wkum+kacWdKSBfSrAHSW3f0zDsIAAABjkkU4wYAAAOD
AECwR0IgeS5We0S8uSe4RvZ02G6EK8a3NGFFu7L3+rDF9JW1ZNTCQC/62e116Hv
F44qkKHrJW+wwLTS4GpA8K0dovx10J60A92AKLjCuRF772tm3447Udm1fPx1UE
1NcrXhssXL0peFnAAABjkkU4Q8AAQDAECwR0IgeYODT71xkFadNo7Y0AKZ1Bggr
KFRAGVRS2dJFV113UC10CGVMBVNIhdbM7SqTPWQ1sk2wYGCC1axnrN1YJUMp
nQB2A551ovdcmmDDOfts1N8/Uusd80COG41pwLH6ZLF1mjnFAAABjkkU408AAAOD
AECwR0IgaDzKfQ0k4qcDexHGqSvAUuLX38Scp09Trdz4Y0Fds0CI0D8XhQgV90Z
qup1KECTVnE1Uof+jjCjCgZgCL8Fom7rTANBkqhk1G9w0BAQWfAA0CAYEAb19t
Kfa1DL1+qeb0wU5kkmwS1CS57YBn6AwkuiThBDJ1MB23E7VzRUZPhRk1M0q
bQMK05vN6A/DB0UcUSiemvFpGaw96oI58kvrSmgun1SmjRbPh4Z1qQE6moSxmF
CFcWT206CymcXIXgLv1Wx1t4Mm55VjBpFy1PvYc2w29cLPLo3H+eHDSXvN1Pe0
f9Ag0B4/na1QJ/N6vAfaceHagVFNXVfz+MmkbrOuQacE12IWo00E6NFRuN7EK1p
0vD9HEM6amRZzcVONC1uhwS2cvQvQ1ANF3jWIX1ctgV206Z7P42F2TmPbecDS
RPF6FW6F1o6A91xoZmCo1vxXtMk2GfH4X9eZ2FwpGgRLB0vQatPctxc2s1omt6yn
GWK5gssmL1xtumRnEW5HFduH6wKE11muMGCF0yPgsgcV3Zd6AogR055K33p3au
UB8oz/p4je00wXfPnGLGMyFBEB9FH7mX628M1MwxW+8sAVI8t1J1TL7W3
-----END CERTIFICATE-----
```

The following is some of the above data in a human-readable format, for informational purposes only:

```
Version: 3 (0x2)
Serial Number: 4b:4e:ae:d4:f5:c5:7f:bd:4a:f2:b2:ec:2e:32:85:78
Issuer: C=US, O=Internet2, CN=InCommon RSA Server CA 2
Validity:
  Not Before: Mar 15 00:00:00 2024 GMT
  Not After : Mar 15 23:59:59 2025 GMT
Subject: C=US, ST=Arizona, O=The University of Arizona, CN=ftp.genome.arizona.edu
```

