

## Aspera Transfer Guide

Created: May 11, 2009; Updated: April 16, 2014.

### Notice

Reference herein to any specific commercial products, process, or service by trade name, trademark, manufacturer, or otherwise, does not necessarily constitute or imply its endorsement, recommendation, or favoring by the United States Government. The views and opinions of authors expressed herein do not necessarily state or reflect those of the United States Government, and shall not be used for advertising or product endorsement purposes.

### Overview

This document provides instructions on the use and installation of Aspera Connect for high throughput file transfer with NCBI. As the sizes of the datasets have increased, we have found that the traditional methods of *ftp* or *http* do not have the performance characteristics needed to support this load of data.

Requirements for large scale data transfer over the internet include high bandwidth, auto checksum, recursive copy, and security based on strong keys. NCBI has chosen to use a product from Aspera, Inc (Emeryville, CA) because of improved data transfer characteristics. FTP and HTTP access will continue to be available and are the default options for users without Aspera installed. Instructions are provided below for investigators to use this data transfer technology. NCBI also is open to using additional products with the appropriate performance characteristics.

### Scope

This document is intended for users transferring large data files to and from NCBI. It applies to the Sequence Read Archive (SRA), dbGaP, and other archives where aspera download is enabled.

### Aspera

#### Aspera Connect

Aspera Connect is software that allows download and upload via a web plugin for popular browsers on machines running Linux, Windows, and Macintosh. The software also includes a command line tool (*ascp*) that allows scripted data transfer. The software client is free for users exchanging data with NCBI.

Download and install Aspera Connect software from: <http://downloads.asperasoft.com/connect2/>

The website's download button will default to the detected operating system of the user's computer. To download for a different OS, click the link to 'See all installers'.

**Please note the Requirements and consult with your network administrator to ensure transfers with aspera will not be blocked.**

Aspera can be installed for individual users. However users of shared machine may want to have the software installed for all users by a system administrator.

## The fasp Protocol

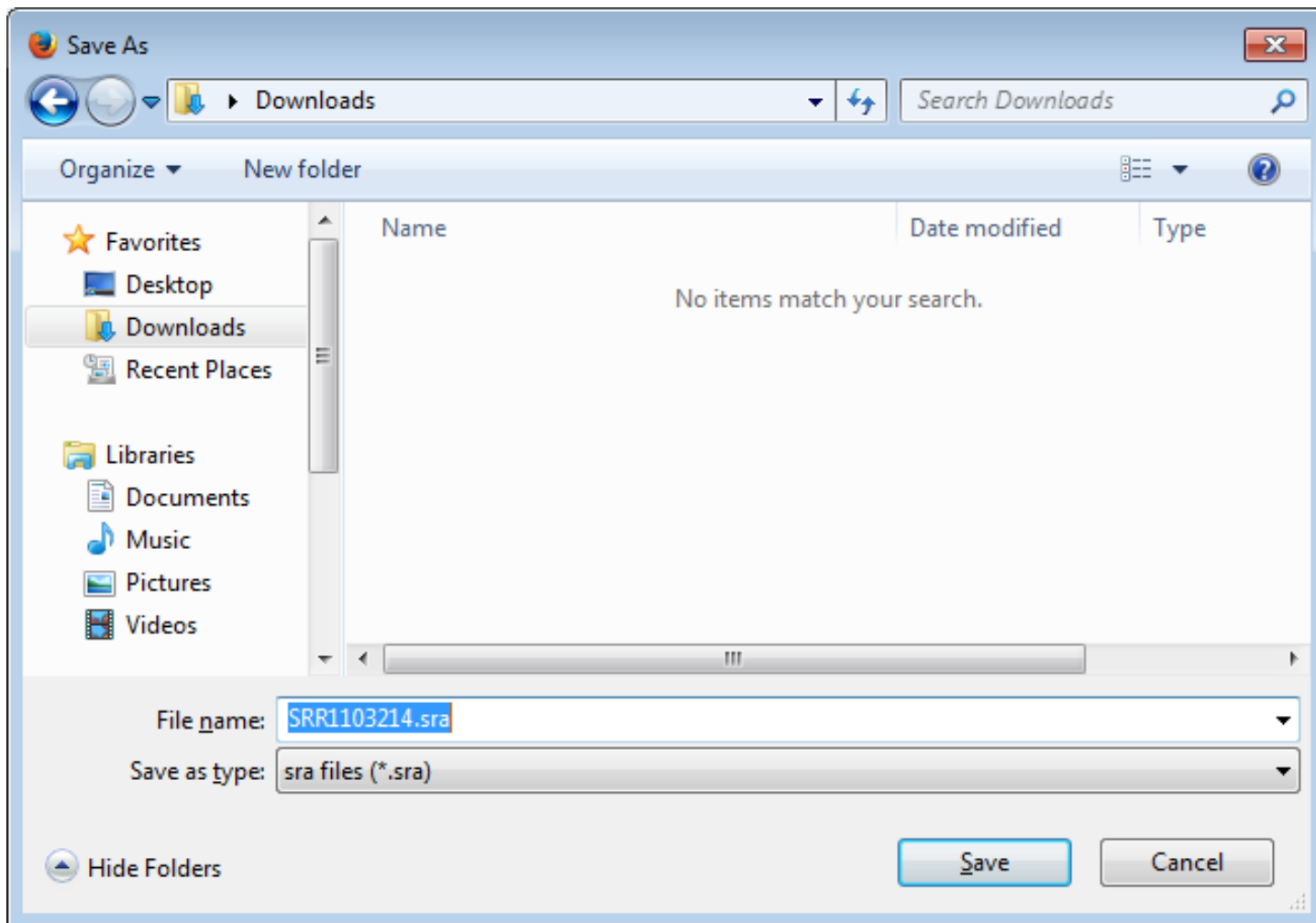
The FASP protocol from Aspera ([www.asperasoft.com](http://www.asperasoft.com)) uses UDP, eliminating the latency issues seen with TCP, and provides bandwidth up to 5 gigabit per second (Gbps) to transfer data. It has a restart capability if data transfer is interrupted midstream and is well behaved, so if there is other data traffic on your network connections, it will back off in order to avoid starving other protocols. We have seen effective throughput up to 800 megabits per second (Mbps) to a single site.

## Downloading Data with Aspera Connect Browser Plugin

Once the plugin has been installed in your browser, you may download files or entire directories from NCBI using Aspera. Example: In your browser window, go to

<http://www.ncbi.nlm.nih.gov/public/?/ftp/sra/sra-instant/reads/ByRun/sra/SRR/SRR292/SRR292241>

Click 'SRR292241.sra' to begin saving the data. You will be prompted to select where the file is to be saved. For example:



You can download full directories or a single file at a time. The Aspera Connect plugin works with Chrome, Internet Explorer (IE), Safari, and FireFox web browsers. In some cases Aspera Connect may create a popup window to get a confirmation for file transfer and this popup window can be hidden behind your current web browser.

## Using ascp to Download by Command Line

The command line program *ascp* is a utility delivered along with the Aspera Connect product.

```
ascp -i <asperaweb_id_dsa.openssh with path> -k1 -Tr -l100m
anonftp@ftp.ncbi.nlm.nih.gov: /<files to transfer> <local destination>
```

- `-i <asperaweb_id_dsa.openssh with path>` = fully qualified path & file name where

this public key file is located. This file is part of Aspera Connect distribution and is usually located in the 'etc' subdirectory.

- `-T` to disable encryption
- `-k 1` enables resume of partial transfers
- `-r` recursive copy
- `-l` (maximum bandwidth of request, try 100M and go up from there)

Experiment with transfers starting at 100 Mbps and working up to 400 Mbps. Select the bandwidth setting that gives good performance with unattended operation.

- `<files(s) to transfer>` = names of files to transfer (including path)
- `<local destination path>` = location to store the downloaded data

## Windows Executable Location

The *ascp* program for Microsoft Windows is located by default in “*C:\Program Files\Aspera\Aspera Connect\bin\ascp.exe*”

## OS X Executable Location

The *ascp* Mac program location is */Applications/Aspera Connect.app/Contents/Resources/ascp*

## Linux Executable Location

The *ascp* Linux program location is */opt/aspera/bin/ascp*

Additional information is available at the Aspera Web site: <http://downloads.asperasoft.com/documentation/>

## Using ascp to Upload by Command Line

In order to use the Aspera upload service you will need to use a **private** SSH key, individual users can contact us at [sra@ncbi.nlm.nih.gov](mailto:sra@ncbi.nlm.nih.gov) to request an Aspera private key.

Upload Command

```
ascp -i <private key file> -T -l 100m <file(s) to transfer>
asp-****@upload.ncbi.nlm.nih.gov: <destination directory>
```

- `-i <private key file >` = fully qualified path & file name of the private SSH key
- `-T` to disable encryption
- `-k 1` enables resume of partial transfers

- -l (maximum bandwidth of request, try 100M and go up from there)

Experiment with transfers starting at 100 Mbps and working up to 400 Mbps. Select the bandwidth setting that gives good performance with unattended operation.

- <files(s) to transfer> = names of files to transfer (including path)
- <destination directory> = deposit location of the uploaded data (typically either 'test' or 'incoming')

For password protected private keys, it is possible to run *ascp* in an autonomous, unattended manner that does not require repeated login. The environmental variable `ASPERA_SCP_PASS` can be used to store the private key path for a scripted series of bulk uploads.

## Key Pairs

SSH keys are used for establishing secure connections to remote computers.

Submitters using a dedicated center account can find instructions for generating a key pair or converting PuTTY format private keys to OpenSSH format in this guide.

<http://www.ncbi.nlm.nih.gov/books/NBK180157/>

## Requirements

### Firewall Requirements

Your local firewall must permit UDP data transfer in both directions on ports 33001-33009 for the following IP ranges:

130.14.\*.\*

165.112.\*.\*

The firewall must also allow ssh traffic outbound to NCBI.

## Troubleshooting

Here are some example commands demonstrating a test download.

Mac OS X:

```
ascp -T -l640M -i "/Applications/Aspera Connect.app/Contents/Resources/asperaweb_id_dsa.openssh" anonftp@ftp.ncbi.nlm.nih.gov:1GB /tmp/
```

Linux:

```
ascp -T -l640M -i /opt/aspera/etc/asperaweb_id_dsa.openssh anonftp@ftp.ncbi.nlm.nih.gov:1GB /tmp/
```

MS Windows:

```
C:\TEMP>"C:\Program Files (x86)\Aspera\Aspera Connect\bin\ascp.exe" -T -l640M -i "C:\Program Files (x86)\Aspera\Aspera Connect\etc\asperaweb_id_dsa.openssh" anonftp@ftp.ncbi.nlm.nih.gov:1GB C:\Temp\
```

For additional assistance, please contact the NCBI Help desk at [info@ncbi.nlm.nih.gov](mailto:info@ncbi.nlm.nih.gov)

When you are about to contact the NCBI Help desk please provide them some basic information like operating system, version of aspera connect, type of disk storage used for transferring files and the type of network connection your organization has to the internet.

If you have a Linux or MacOS X operating system you may run these commands and show us their output:

```
curl -o /dev/null ftp://ftp.ncbi.nlm.nih.gov/1GB
curl -o /dev/null http://www.ncbi.nlm.nih.gov/staff/beloslyu/large.tar
traceroute ftp.ncbi.nlm.nih.gov
```

First two commands download a 1GB file from NCBI using ftp and http protocols, the content is dumped to /dev/null. The third command will let us see the latency in your internet connection and possible congestions on the way to NCBI.

Another possibility is to make some test downloads from Aspera's demo server, for Linux the command line is:

```
env ASPERA_SCP_PASS=demoaspera ascp -L- -T -l100m aspera@demo.asperasoft.com:aspera-test-dir-large/1GB /tmp/
```

Aspera Connect is a commercial product and program specific support is available from the manufacturer at <http://asperasoft.com/support/>

The currently up-to-date documentation for ascp can be found at <http://downloads.asperasoft.com/en/documentation/8>