

# Data Import to the Health Data Platform BC – Trusted Analysis Environment

## Contents

Data Import to the Health Data Platform – Trusted Analysis Environment.....	1
Introduction for Data Contributors .....	2
Process Overview for Data Contributors.....	2
Preparing Your Data Files for TAE-BC.....	3
File Naming Standards.....	3
Token Definitions .....	3
Data Filename Standard .....	4
Control File Standard .....	4
Metadata Filename Standard .....	5
Preparing the Metadata Information for Your Data .....	5
Preparing Your Data File(s) .....	6
Preparing Your Control File(s) .....	7
Uploading Your Data to TAE-BC SFTP .....	8
Step 1: Notify TAE-BC Support Team of Readiness .....	9
Step 2: Encrypt and Compress Your File(s) .....	9
Step 3: Upload Your Files to the TAE-BC SFTP .....	13
TAE-BC SFTP Configuration .....	13
Step 4 -Notify TAE-BC Team File Transfer is Complete.....	16
Need Support? .....	16

## Introduction for Data Contributors

The Trusted Analysis Environment (TAE-BC) integrates data holdings across organizations enabling access to safe and governed services that support person centric longitudinal analysis to further public good for British Columbians.

TAE-BC enables collaborative, governed, and safe analytical services in the Province of British Columbia. Data provisioned on behalf of the organizations is accessible within a secure, internet-free environment. All information going in and out of this secure space undergoes human review, guided by comprehensive agreements that define the relationship and outline roles and responsibilities of all parties involved. We offer:

- **Data Access Management:** Complete handling of the data access request lifecycle, from consumer engagement to Health Data Council approvals
- **Resource Provision:** Central team for data assembly, regular collection, and refreshes
- **Data Import:** Seamless importation of complete datasets into a fully secure environment
- **Data Agility:** Enhanced data consumer agility within approved research parameters
- **Request Efficiency:** Reduced need for the organizations to service individual research requests
- **Streamlined Approval:** Quick approval process following the 5 Safes Framework for multi-organization requests

Your data is actively managed and stored. We consistently update metadata to ensure accuracy and regularly refresh the datasets to keep information current and relevant.

The data you provide is utilized in a range of ongoing research projects, driving innovation and enhancing understanding across multiple fields.

## Process Overview for Data Contributors

Now that you're ready to share your data with TAE-BC, we've outlined some quick and easy steps to transfer your data securely:

- Preparing your data and metadata according to TAE-BC standards
- Encrypting data according to our standards
- Uploading your data to the SFTP server following the instructions outlined in this document

- Notifying the TAE-BC team that the data has been uploaded
- TAE-BC team will ensure there's a human review/quality checks on the data and will work with data contributors if there are any questions
- TAE-BC team will ensure that the data is ready to share with approved projects

For any questions, reach out to TAE-BC Support at [TAEDATAC@Victoria1.gov.bc.ca](mailto:TAEDATAC@Victoria1.gov.bc.ca)

## Preparing Your Data Files for TAE-BC

This section describes the general steps and naming standards for preparing the following data files:

- Metadata Files
- Data Files
- Control Files

### File and Data Naming Standards

TAE-BC requires that a file naming standard is applied to all incoming files to ease the governance and automation of the data pipeline.

#### Special Characters

Some special characters cannot be included in your data file

- '-' (hyphen) is reserved for separating tokens
- '.' (decimal/period) is reserved for file type delimiter
- '<' and '>' are reserved for new line characters or html tags

#### Column names

- Do not begin with a number
- Do not end with '\_e'

#### Token Definitions

- [DC Label]: A short label, typically abbreviated title of your organization. **Once you choose a DC label, please ensure the same label is used for all future transfers. This label should not be more than 8 characters long.**
- [Data Holding Label]: A short label of data holding or data group label. This is typically an abbreviated title of a logical collection of data objects that are expected

to be transferred together. **Once you choose a data holding label, please ensure the same label is used for all future transfers of the same data holding. This label should not be more than 16 characters long.**

- [Data object name]: This is a unique name that identifies file contents, exclusive of any versioning information. **This label should not be more than 64 characters long.**
- [Data shard label]: If the data object is partitioned, use this label to differentiate the object's shards. Typically, this label is a year or a range of years. Note: put 'NA' if the object is not a shard.
- [DC Transfer Date]: Date on which the data was transferred to TAE-BC in YYYYMMDD format (i.e., 20240121)
- [Content Type]: Indicate whether the file is data, 'meta' or 'CTL' (control file). **For data files, please omit this token.**
- Tokens can include underscores "\_". Hyphens "-" are reserved for separating tokens.

Please let the TAE-BC Support at [TAEDATAC@Victoria1.gov.bc.ca](mailto:TAEDATAC@Victoria1.gov.bc.ca) know if there are concerns with meeting any requirements relating to the file naming standards.

### Data Filename Standard

- **Standard:** [DC Label]-[Data Holding Label]-[Data object name]-[Data shard label]-[DC Transfer Date].csv
- **Examples:**
  - HI-CommonDims-AGE-NA-202401024.csv
  - HI-CommonDims-LHA\_GEOG-NA-202401024.csv
  - HI-DAD-DAD\_CORE-2024-20250121.csv

### Control File Standard

- **Standard:** [DC Label]-[Data Holding Label]-[Data object name]-[Data shard label]-[DC Transfer Date]-CTL.ctl
- **Examples:**
  - HI-CommonDims-AGE-NA-202401024-CTL.ctl
  - HI-CommonDims-LHA\_GEOG-NA-202401024-CTL.ctl
  - HI-DAD-DAD\_CORE-2024-20250121-CTL.ctl

## Metadata Filename Standard

- **Standard:** [DC Label]-[Data Holding Label]-[Data object name]-NA-[DC Transfer Date]-meta.csv
- **Note:** All sharded objects should have the same metadata so only one metadata file is expected for all shards. Please use the same naming structure as you would for a shard but put 'NA' into the [Data shard label] token position.
- **Examples:**
  - HI-CommonDims-AGE-NA-202401024-meta.csv
  - HI-CommonDims-LHA\_GEOG-NA-202401024-meta.csv
  - HI-DAD-DAD\_CORE-NA-20250121-meta.csv <- note 'NA' for a typically sharded object.

## Preparing the Metadata Information for Your Data

File Naming Standard	[DC Label]-[Data Holding Label]-[Data object name]-NA-[DC Transfer Date]-meta.csv
File Format	.csv
Other Requirements	<ul style="list-style-type: none"> <li>• <b>Please create one '-meta' file for each data object file that you are transferring</b> (with exception of shards, where a single '-meta' object is sufficient for all shards)</li> <li>• Please let us know if there are concerns with meeting any of the metadata requirements, or if metadata is already in a standard, but different, format (e.g. Mauro Metadata)</li> <li>• Column names should not begin with a number</li> <li>• Column names should not end in '_e'</li> </ul>

The table below shows the mandatory metadata information required:

Required Metadata Information			
field_name	data_type	description	identifier_flag (Direct or In Direct)

To apply privacy protection rules to the data being imported to your project, the TAE-BC Technical Support team needs to know which columns are direct identifiers, and indirect identifiers.

Direct identifiers, including those with null or proxy values, will be encrypted to facilitate linking of your cohort with your TAE-BC project data in the platform. Please include additional descriptive columns to flag these values for use in your analysis, if appropriate.

Ensure that first and last names have been excluded from your file prior to transferring.

- **Direct Identifiers** are variables that can directly identify individuals, either individually or in combination. e.g., name, email address, IP address, home address, personal health number (PHN), social insurance number (SIN).
- **Indirect Identifiers** are variables that don't directly identify an individual but can be used for indirect re-identification. Whether a variable is an indirect identifier depends on the nature of the data.

Example of a populated Metadata Information Table:

File name: HLTHORG-HLTHDH-HospVisit-NA-20250121-meta.csv			
field_name	data_type	description	identifier_flag
PHN	varchar(50)	Personal Health Number	Direct Identifier
Name	varchar(50)	First and Last Name	Direct Identifier
Diag_Code	int	Code representing diagnosis	
FSA_location	varchar(250)	First three digits of postal code	Indirect Identifier
home_address	varchar(250)	Home address	Direct Identifier
attend_time	float	Time spent on patient	
ICD_10_code	int	ICD 10 Code	
PHN	varchar(50)	Personal Health Number	Direct Identifier
Mother Name	varchar(50)	First and Last Name of Mother	Direct Identifier
Diag_Code	int	Code representing diagnosis	

## Preparing Your Data File(s)

File Naming Standard	[DC Label]-[Data Holding Label]-[Data object name]-[Data shard label]-[DC Transfer Date].csv
File Format	.csv

<b>Other Requirements</b>	<ul style="list-style-type: none"> <li>• Ensure your data is extracted with UTF-8 encoding to prevent the generation or inclusion of special characters that are not accepted by the TAE-BC.</li> <li>• Column names should not begin with a number</li> <li>• Column names should not end in ‘_e’</li> </ul>
---------------------------	---

## Preparing Your Control File(s)

<b>File Naming Standard</b>	[DC Label]-[Data Holding Label]-[Data object name]-NA-[DC Transfer Date]-CTL.ctl
<b>File Format</b>	.ctl
<b>Other Requirements</b>	<ul style="list-style-type: none"> <li>• See information below</li> </ul>

Control files are used to validate data integrity throughout the data pipeline.

- Please create one ‘-CTL’ file for each data object file that you are transferring
- Mandatory control requirement consists of:
  - Row counts
- Other desirable control requirements include:
  - Column counts
  - Distinct counts for categorical columns – **Important: Please ensure to exclude NULL or “” (blank values) in the distinct count calculation.**
  - Sums for numeric columns
- Please use the following JSON format within your .ctl file:

```
{
  "rowCount" : <row_count_value>,
  "columnCount" : <column_count_value>,
  "column" :
    [
      {
        "column" : <column_name>
        "type" : "Distinct Values Count"
        "value" : <distinct_count_value>
      },
    ]
}
```

```

        {
            "column" : <column_name>
            "type" : "Sum"
            "value" : <sum_value>
        }
    ]
}

```

#### Sample JSON Format:

```

{
  "rowCount" : 3036,
  "columnCount" : 24,
  "columns" :
  [
    {
      "column" : "Col_1",
      "type" : "Distinct Values Count",
      "value" : 3036
    },
    {
      "column" : "Col_2",
      "type" : "Distinct Values Count",
      "value" : 1033
    }
  ]
}

```

## Uploading Your Data to TAE-BC SFTP

Using a secure file transfer service ensures sensitive data is protected through encryption and controlled access. It supports compliance with regulatory standards and provides detailed audit logs for traceability. The following sections outline the steps to securely share your organizational data.



## Step 1: Notify TAE-BC Support Team of Readiness

Each Data Contributor who needs to share data will require a TAE-BC SFTP account to ensure all exchanges are encrypted and compliant. Once your data and associated data files are nearly ready, email the TAE-BC Support team ([TAEDATAC@Victoria1.gov.bc.ca](mailto:TAEDATAC@Victoria1.gov.bc.ca)) with the following information:

- The Public IP address of the computer that will be sending data
- Primary TAE-BC SFTP Contributors (for each individual sharing data)
  - Full Name
  - Email
  - Organization

A public IP address is needed for the TAE-BC SFTP to be accessible to Data Contributors from outside our private network. How to find your public IP address:

1. Confirm you are connected to your organizational network
2. Go to this [link](#)
3. Your public IP address will be displayed on the website as IPv4

Once all the required details are sent to the TAE-BC team, a TAE-BC SFTP username and password will be created and assigned to the folders. These credentials will be shared by the TAE-BC Support team and allow you to test your connection and share data files.

## Step 2: Encrypt and Compress Your File(s)

To protect sensitive information, it's important to encrypt files before sharing them. The following steps will guide you through encrypting your files securely. TAE-BC has the following encryption requirements:

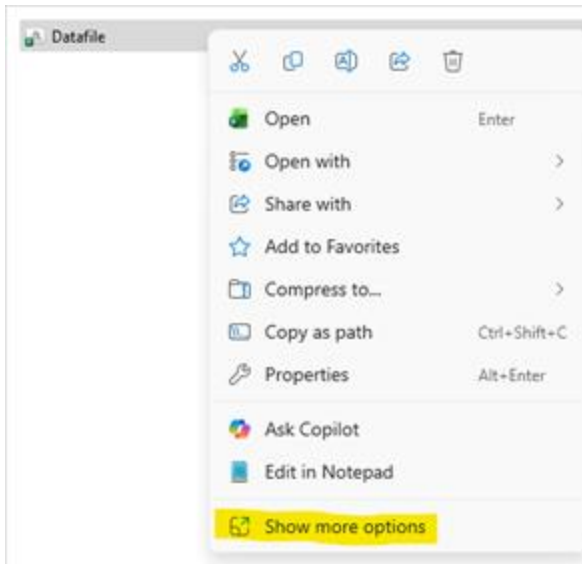
- Data Files – Encryption is required
- Metadata and Control Files – Encryption is not required (optional)

*If you are submitting a refresh of previously contributed data, please use the same encryption pass phrase as your first transfer, and for all subsequent transfers.*

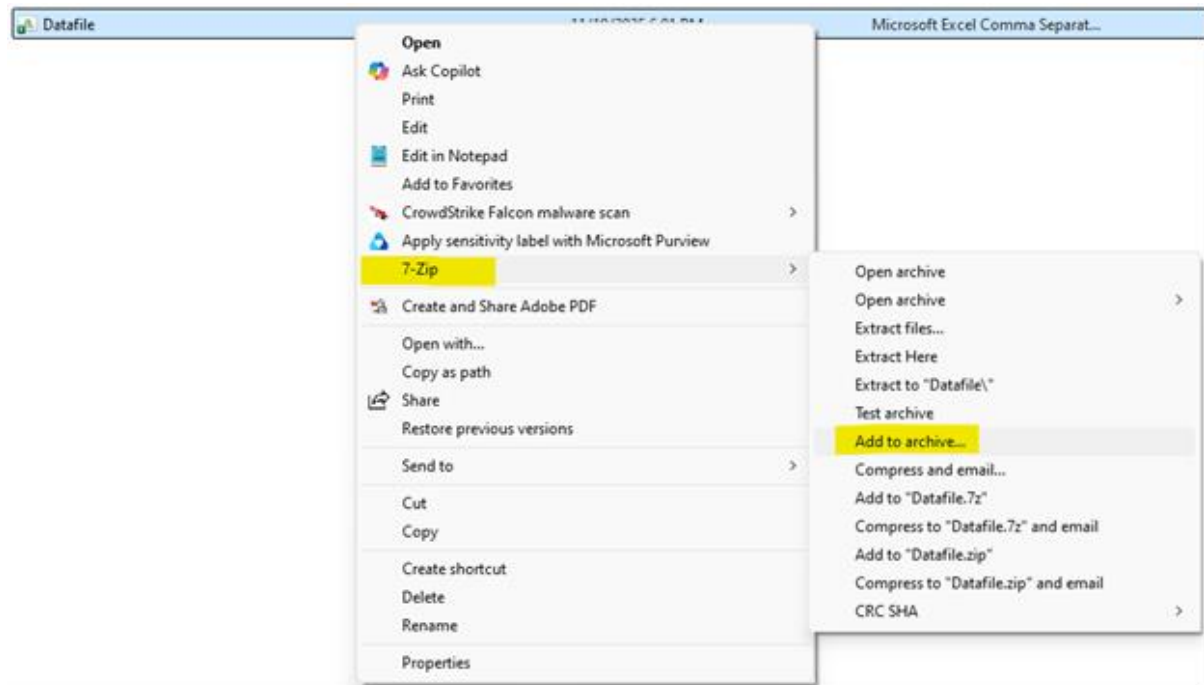
To get started with encrypting your files, follow the steps below. This will help ensure your data remains secure during transfer and only accessible to authorized recipients.

1. Confirm or download [7-Zip](#) on your machine

- a. You can confirm 7-Zip is on your machine by going to the Windows start menu and search for 7-Zip. If 7-Zip appears in the search results you have this downloaded.
  - b. To download 7-Zip navigate to the official 7-zip website - <https://www.7-zip.org/> and follow the steps to download
2. Once you confirm 7-Zip is on your machine, go to the data file you want to compress and right-click, select Show more options

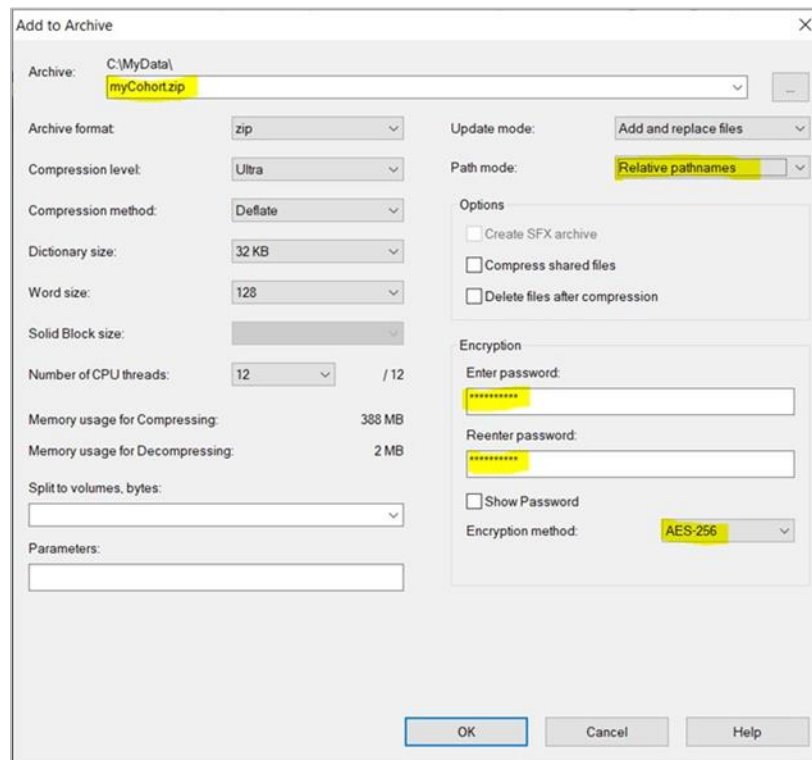


3. Then select 7-Zip and add to archive from the 7-zip submenu.

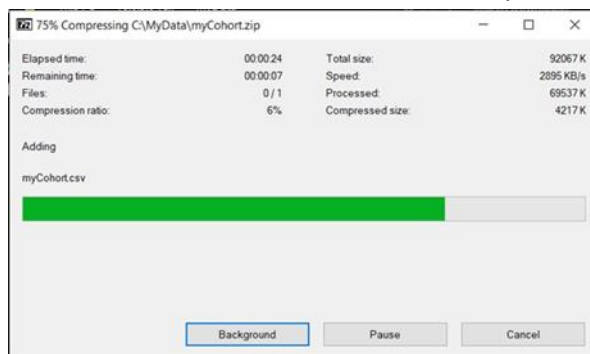


4. Ensure the following details are entered as shown in screenshot below:
  - a. Enter the file name
  - b. Next to **Path mode** ensure **Relative pathnames** is selected
  - c. Under the Encryption section:
    - i. Enter a strong password under **Enter Password** and **Re-enter Password** and save it for future use. This password should be used for all the files uploaded to TAE-BC SFTP until otherwise notified by the TAE-BC team
    - ii. For Encryption method, select **AES-256**

d. Leave the rest of the fields as default



5. Press OK and wait for the file to compress



6. Once the file is compressed you will see a new zip file. This is the file you will upload to the TAE-BC SFTP, in addition to your metadata information and control file(s).

Now that your data is compressed, you're ready to upload it to the TAE-BC SFTP for safe and encrypted transmission.

## Step 3: Upload Your Files to the TAE-BC SFTP

### TAE-BC SFTP Configuration

As TAE-BC data involves Personal Identifiable Information, it is important that the data files are encrypted and transmitted safely through a Secure File Transfer Protocols (SFTP) server. Secure file transfer servers establish a protected connection for data transfers, ensuring a high level of security and safeguarding the data throughout the transfer process.

The steps below outline how you can upload your data, metadata, and control files to the TAE-BC SFTP. Please note, depending on your company policies, you may require Administrator privileges on your desktop to complete the installation.

#### 1. Download and Install an SFTP Client

There are a few SFTP Client options to choose from based on your preferences.

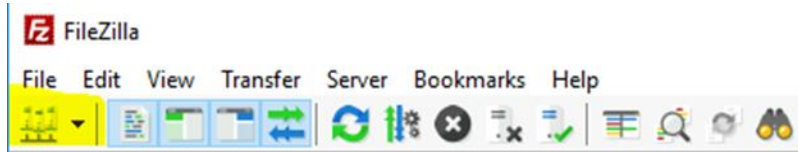
Note: If you are a BC Ministry employee, the software required to upload your files can be found in the Ministry Company Portal (formerly “Software Center”). From there you can find the application (FileZilla, WinSCP, or Cyberduck) and install it on your desktop.

- a. The other option is to visit <https://filezilla-project.org> and click on the "Download" option in the left-hand navigation menu. Follow the prompts to download and install the appropriate version for your operating system.



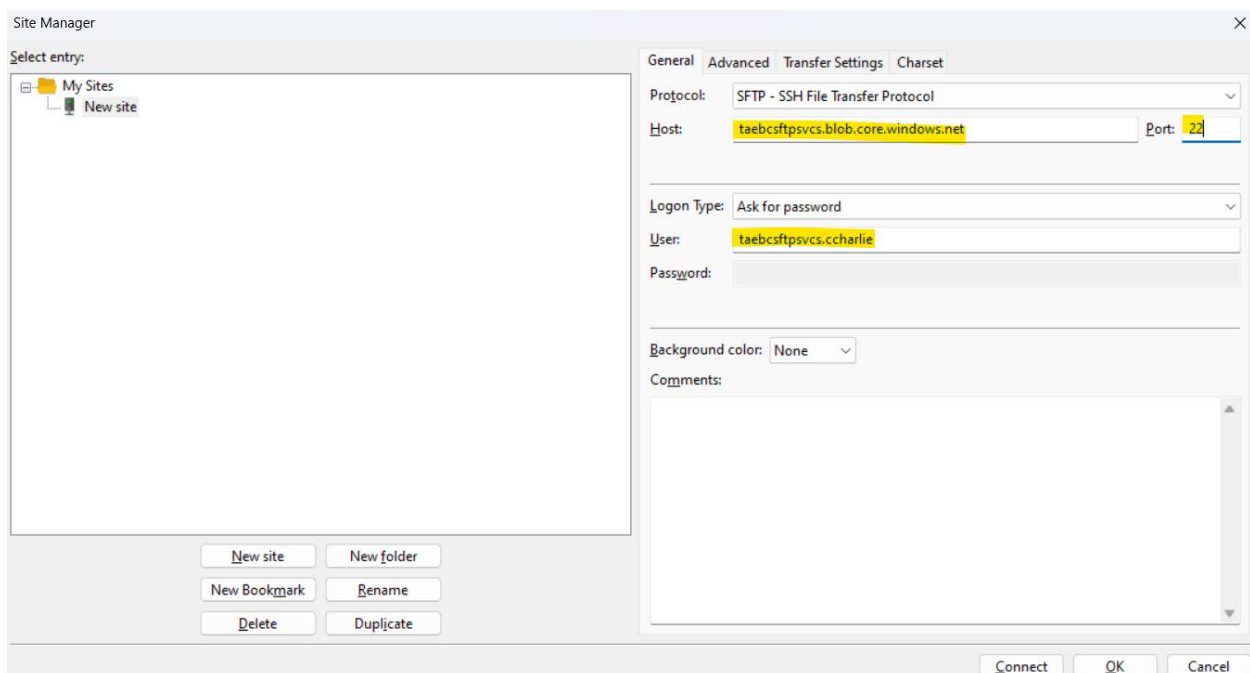
The steps below are shown using FileZilla, but the process is similar for WinSCP and CyberDuck as well.

- Open FileZilla and select the first button on the top lefthand side to start a new connection.



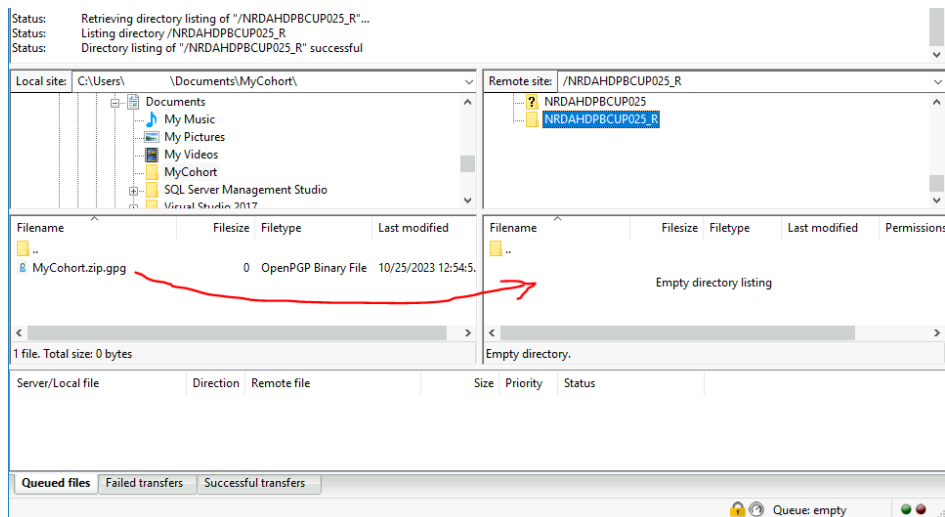
- Next, you will be prompted to fill in the connection information as follows:

TAE-BC SFTP Connection Information	
<b>Protocol</b>	SFTP
<b>Host</b>	taebsftpsvcs.blob.core.windows.net
<b>User</b>	taebsftpsvcs.{replace with the username, provided by TAE-BC team} For e.g. taebsftpsvcs.ccharlie
<b>Password</b>	Replace with the password provided by TAE-BC team
<b>Port</b>	22

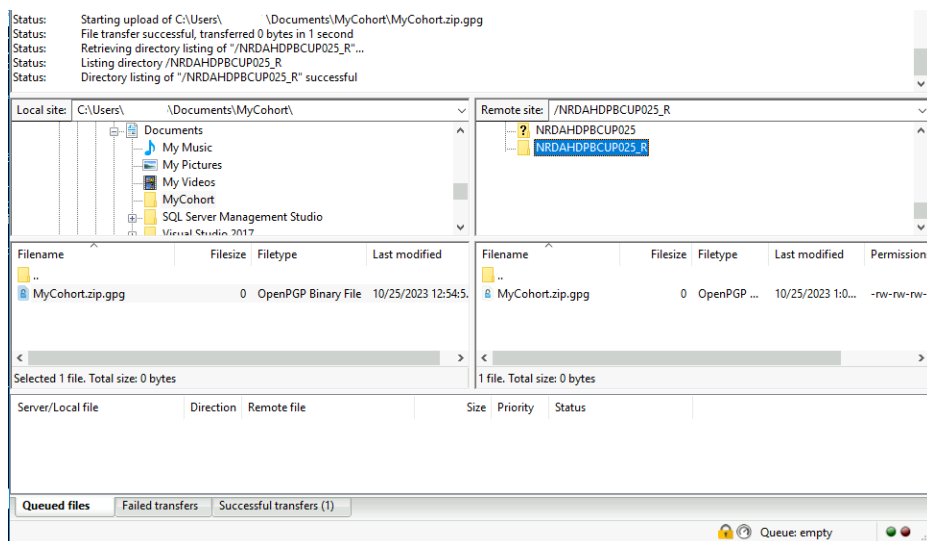


- Once all of the information is input you can select Connect. When you click **Connect**, it will ask you for password. Enter the password and click **OK** to connect.
- If you've connected successfully the "Remote site" panel to the right will be populated with the folders you have access to in the TAE-BC SFTP.

- In the “Local site” pane on the left side of the page, you can navigate to the files or folders on your computer that contains your encrypted data files. Once you find the files you want to share you can click and drag the folders or files from the **Local site** to the **Remote site**. See screenshot below.



- To confirm your file transfer was successful you will see the file appear in the ‘Remote site’ on the righthand side of the window and the ‘Successful transfers’ tab at the bottom of the window will show details about the transfer.



## Step 4: Notify TAE-BC Team File Transfer is Complete

- Notify the TAE-BC Support team ([TAEDATAC@Victoria1.gov.bc.ca](mailto:TAEDATAC@Victoria1.gov.bc.ca)) that the file(s) have been uploaded for the TAE-BC Technical Support team to pick up—**please include your encryption password** (refer to Step 3 in the Encrypt and Compress your files section) for file retrieval.
- The TAE-BC Support team will inform you once the data load process is successfully completed.

## Need Support?

Technical Support or Questions	Contact TAE-BC Support ( <a href="mailto:TAEDATAC@Victoria1.gov.bc.ca">TAEDATAC@Victoria1.gov.bc.ca</a> )
--------------------------------	--