Ambient Water Data Exchange (awX) System User Guide

awX Pre-release Version 0.3.10

DRAFT - **March 17, 2021**

Connecticut Department of Energy and Environmental Protection
Monitoring and Assessment Program
Bureau of Water Protection and Land Reuse
79 Elm Street
Hartford, CT 06106

Table of Contents

_____

1.0 Glossary

Placeholder. Add a glossary of common terms used to describe the system.

_____

2.0 System Overview

This project builds on previous projects funded by EPA to establish an Exchange Network data flow (EPA Grant OS-83325101-2).   Our goal is to significantly increase the data that CT DEEP flows to EPA's WQX system to support easier and more transparent public access to CT DEEP's water quality data. We are replacing our current data system with a new system that meets the current WQX schema requirements and provides an automated transfer mechanism of data from our local system to WQX. To facilitate this, we are developing a system that meets WQX 3.0 schema requirements and flows data to WQX using the WQX Web node and Application Programming Interface (API). In addition, new modules will be added to the system to flow new data elements and parameters to WQX.
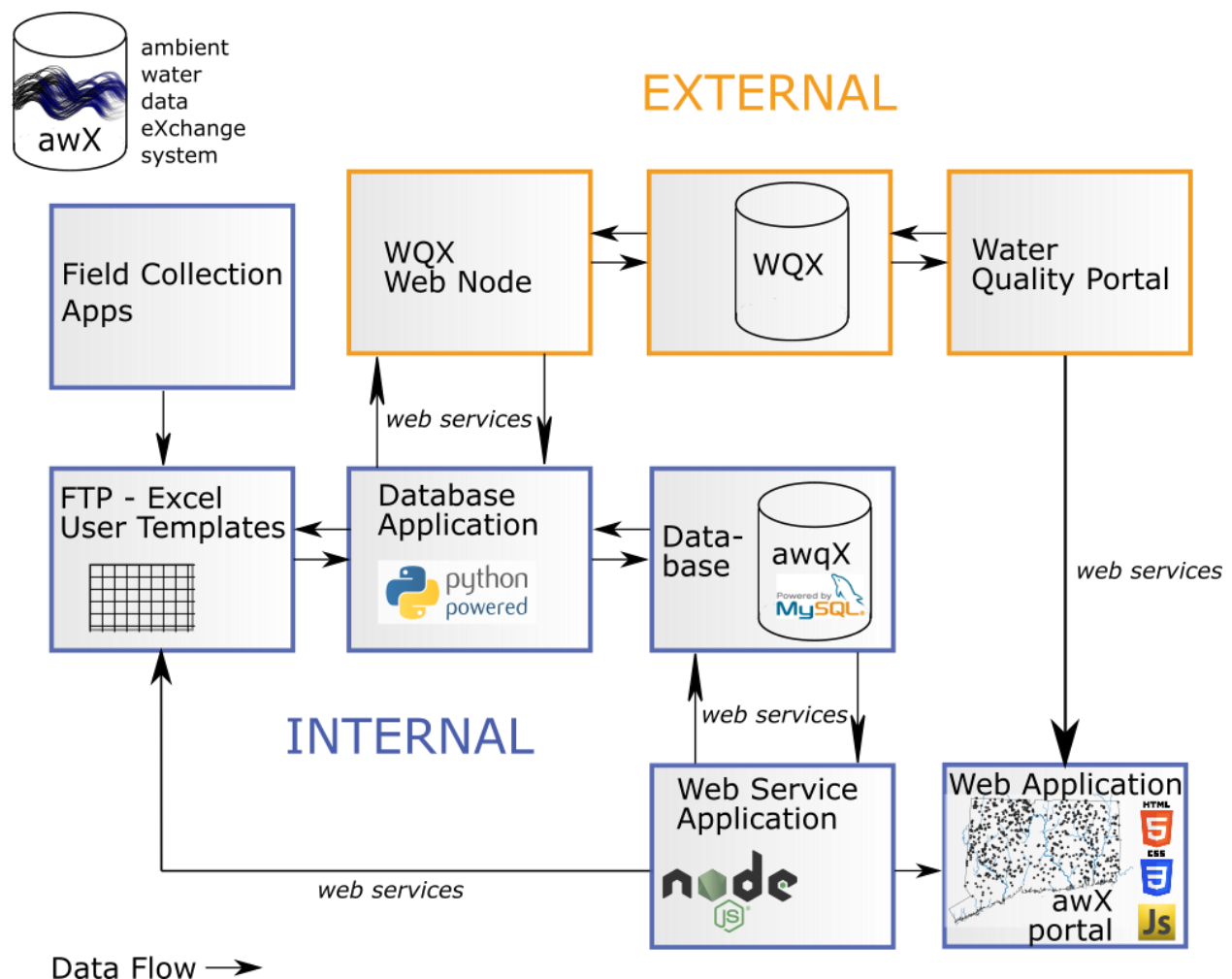

Figure 1. A high level overview of the awX system and how it feeds into WQX and the water quality portal.

_____

The new CT DEEP Ambient Water Data Exchange system automates data uploads to WQX from a locally maintained database system (MySQL) and make web service calls using Python. It also provides users at CT DEEP tools to upload stations, project and activities to the internal system that will flow to WQX after sufficient QA/QC checks by project leads.
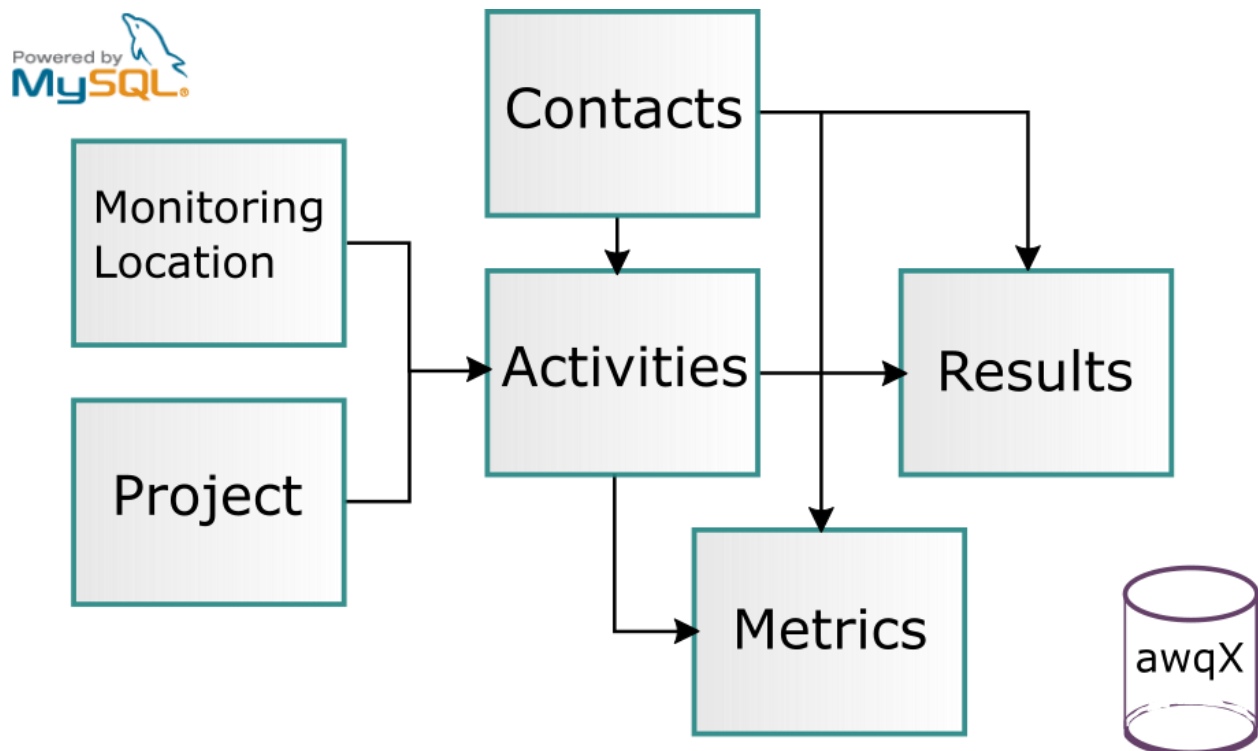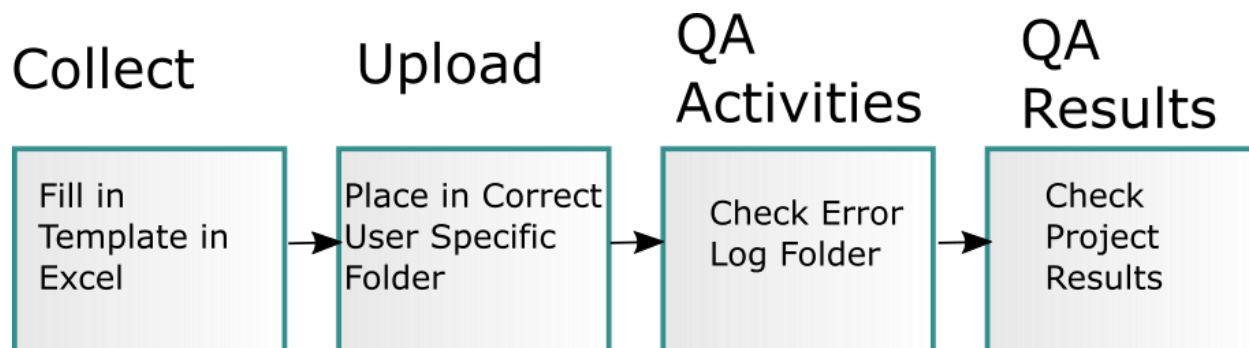
2.1 Overview of awX – awqX database schema

Figure 2. Simplified overview of awqX schema. awqX stores water quality data.

_____

2.2 Overview of User Workflow

Figure 3. Overview of user workflow

_____

3.0 awX portal

3.1 Stations Maps

The Stations map provides a way to see all of the sites in the database. It also provides a way to extract the needed information to create a new stations location using the digital mapping source. When you first create a station it will appear in red for up to a week after it is create, providing another way for users to check newly created stations.

_____

4.0 Connecting to and Navigating the FTP

Open Windows File Explorer type in the file path ftp://sdc-epafiling/ and when prompted for a password type in your Windows authentication (network login) user name and password:
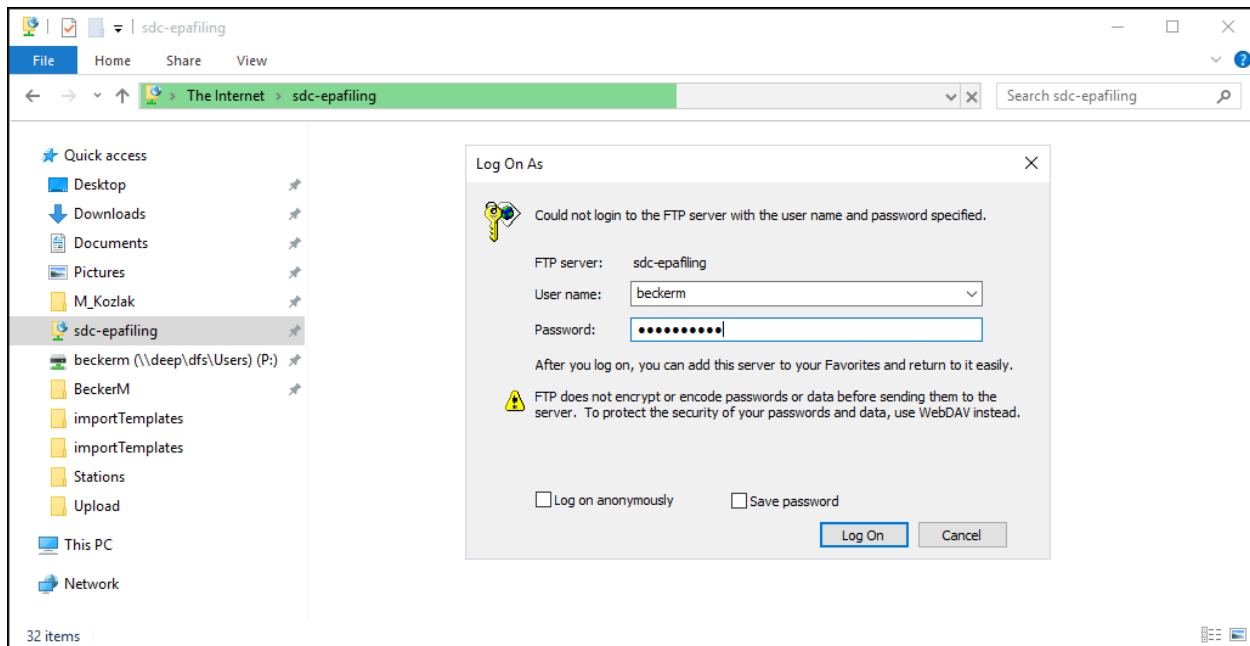

Figure 4. Example of connecting to the ftp in Windows File Explorer.

Only authorized users will be able to access the ftp. If you require access contact the system administrator.

The **importTemplates** folder contains templates which are constrained spreadsheets (user forms) that can be filled out to insert data to the database. Make a copy of a particular template or the entire folder onto a local drive.
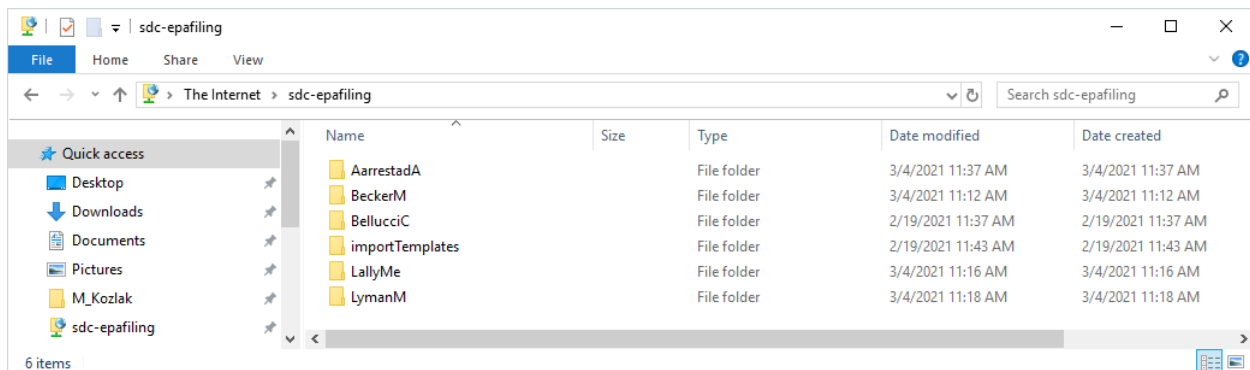

Figure 5. Example of main folder display after logging in to the ftp

_____

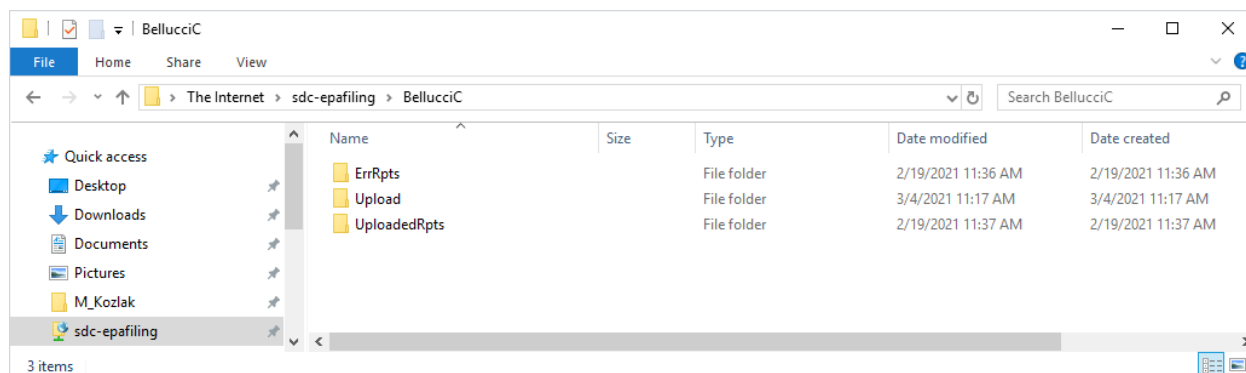Your personal folder includes three folders, **ErrRpts**, **Upload**, and **UploadedRpts**:

Figure 6. Subfolders contained in each user's personal folder include ErrRpts, Upload and UploadedRpts

ErrRpts – Contains rows of data from attempted inserts that were unsuccessful along with the database error that caused the insert issue.

Upload – Contains three sub-folders: **Stations**, **Projects** and **Activities** corresponding to different type of data inserts. Completed data templates are placed in one of these folders. awX will search for data in each of these folders every 5 minutes.

UploadedRpts – Filled in templates placed in the **Upload** folder are moved in their entirety to the **UploadedRpts** folder when awX attempts to upload the data.

FTP tips:

N.B. The folder location you are accessing on the FTP is on another server. You cannot open files directly from the FTP location. The file needs to be copied to a local or network drive first before opening.

It may be helpful to 'pin' the ftp location in Windows File Explorer to the quick access location in the navigation bar.

_____

5.0 Data Insert

The following sections will describe the process of inserting data to the ambient water quality data exchange database.

5.1 Templates

5.1.1 Navigating the Stations Upload Template

Copy the *Stations_Upload_Template.xlsx* file from the **importTemplates** folder on the ftp and open in Excel. This template is used to enter in new Stations data. The *Stations_Upload_Template.xlsx* file includes three worksheets, *StationsInsert*, *fk* and *FieldDescription*:
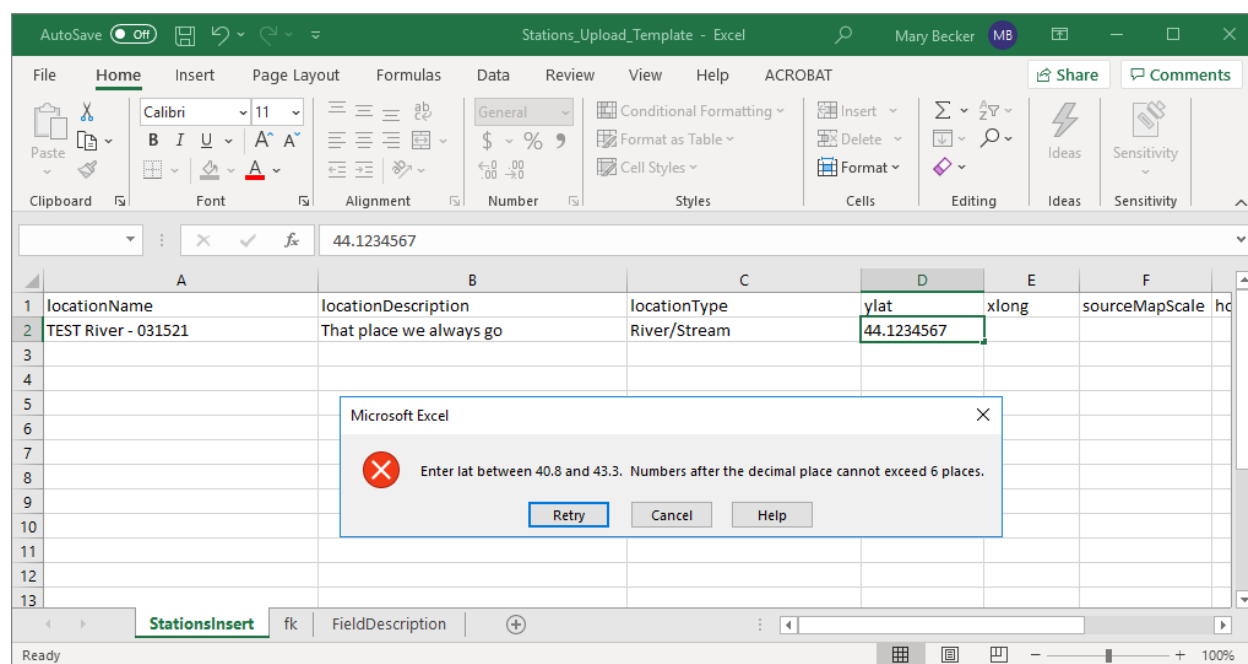

Figure 7. The Stations_Upload_Template.xlsx workbook contains three worksheets: StationsInsert, fk and FieldDescription. StationsInsert is where users enter new data to insert into the database. Some fields are constrained based on data validations rules to help format the data properly for data insert to the database.

StationsInsert – This worksheet contains the attribute fields needed to create a new awX station. This is where users enter data to create new stations. Certain attribute fields are constrained by either foreign key (i.e. pick list) values from the database or by check constraints from the database. For example, the ylat (latitude) and xlong (longitude) fields must be a number value represented a location in or around Connecticut and the numbers after the decimal place cannot exceed 6 places. The data validation rules in Excel are used to help format the data properly for data insert to the database and prevent upload errors.

_____

N.B. The *StationsInsert* worksheet must be the first worksheet in the worksheet tabs for proper data upload. If it is not set as the first tab, the user will receive an error.

fk - This worksheet stands for foreign key. It contains lists of data used to generate pick lists used in the *StationsInsert* worksheet. A foreign key is a field in a database table that is constrained based on another table in the database. In many cases these constrained values are based on EPA WQX domain tables that require that only certain values are entered into a field. For example, the stateCd (State Code) value must be entered a certain way (e.g. CT) to avoid multiple ways of typing in Connecticut. This table is locked for editing. If you believe any field values need updating please contact the system administrator.

FieldDescription – This worksheet provides a brief description of each of the fields in the Stations Upload Template and notes on entering the data. It also notes if a particular field is required for database entry. If a required field is not entered, upon upload, the user will receive an error.

Table 1. Fields, field descriptions and required entries in the Stations Upload Template.

Field	Description	fk	Required
locationName	Name of the waterbody		YES
locationDescription	A brief description of the site location		YES
locationType	Type of waterbody	YES	YES
ylat	Latitude - Decimal Degree - Limited to 6 digits after the decimal. The latitude / longitude combination identifying a site must be unique. Users cannot enter in a duplicate site at the same lat/long location.		YES
xlong	Longitude - Decimal degree - Limited to 6 digits after the decimal. The latitude / longitude combination identifying a site must be unique. Users cannot enter in a duplicate site at the same lat/long location.		YES
sourceMapScale	The scale of the map at which the map was drawn. The hydro USGS layers on the awx portal stations map were drawn at a 1:24000 scale. If unknown or GPS enter 0.		YES
horizCollectMethod	The name of the method used to determine the lat / long values.	YES	YES
horizRefDatum	The name describing the reference datum to determine the lat / long values. Most web maps included the awx stations maps and google map use a WGS84 datum.	YES	YES
stateCd	Name of the state where the site is located.	YES	YES
munName	Name of the municipality where the site is located. Can be unknown, particularly for LIS sites.	YES	YES
subBasin	Number of the subregional basin where the site is located. Use 0 for LIS.	YES	YES
adbSegID	The 305b assessment database segment unique identifier		
hydroID	A unique hydrologic identifier used to associate a site with upstream drainage calculations such as total drainage area		

_____

5.2 Upload Data

Users should save a copy of the template file, such as *Stations_Upload_Template.xlsx* file under a different name prior to entering in data.

N.B. The name of the file needs to begin with the type of data being uploaded and the file type must be an .xlsx file.

For example, for Stations data the file must begin *Stations* (case sensitive) and needs to be an .xlsx file.

Example of file naming conventions: ***Stations_TEST_data.xlsx,***

Projects_TEST_data.xlsx

ChemActvites_TEST_data.xlsx

If these naming conventions are not followed the file will not be recognized for by the insert application and the data will not be uploaded into the database. For example *Stations_TEST_data.xlsx* would be recognized by the insert application and would attempt to upload the data, however *TEST_file.xlsx* or *Stations_Test_data.csv* or *station_TEST_data.xlsx* would not be recognized or uploaded.

Also, if you place a Projects file in the **Stations** Upload Folder, an attempt to upload will not occur.

After users enter data into the template as described above, a copy of the file should be placed on the ftp site (ftp://sdc-epafiling/) in the user's personal folder->**Upload** Folder->Data Activity Name Folder (either **Stations**, **Projects**, or **Activities** folder).
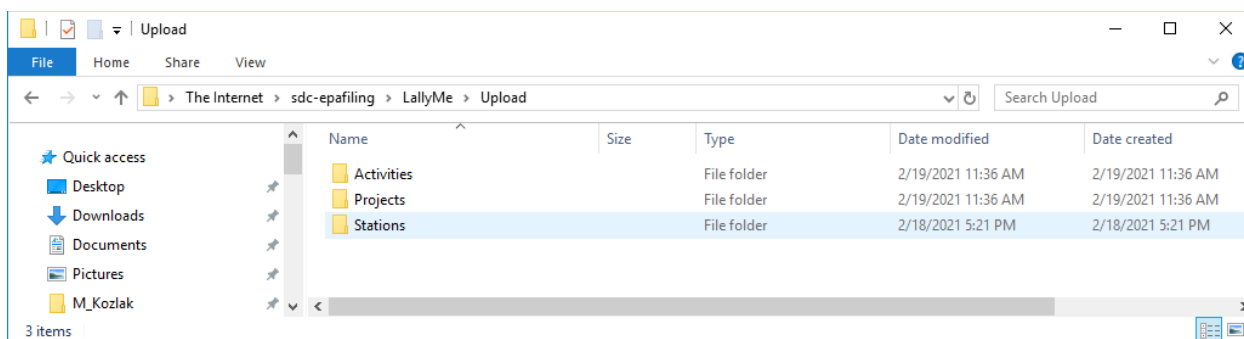

Figure 8. Place data template in it's corresponding data type folder, either Stations, Projects or Activities.
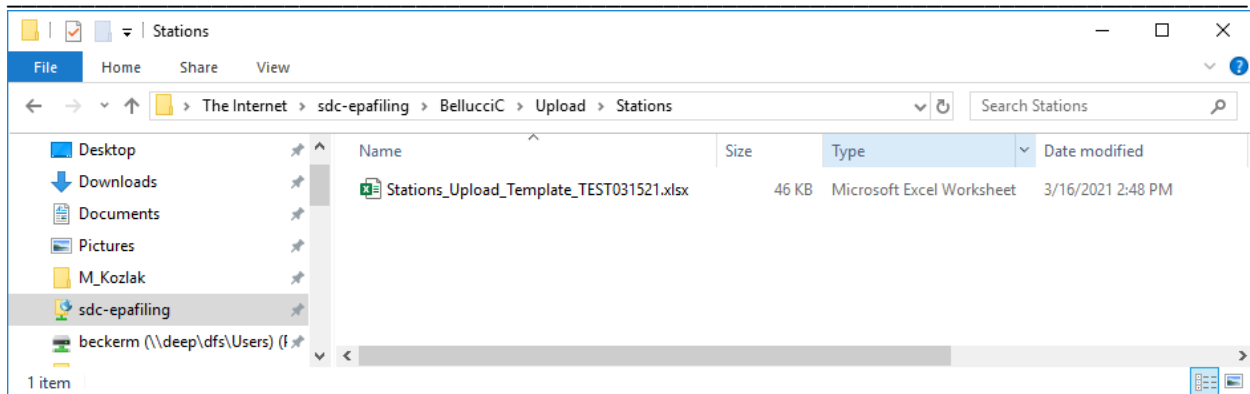
Figure 9. Example of where to place a completed stations template to upload new stations to the database. Once placed in the appropriate folder on the ftp site, the database insert application will attempt to upload the new information to the database within a 5 minute timeframe.

N.B. The database insert application will attempt to upload the new data within 5 minutes.

5.3 Check for Successful Upload

After 5 minutes check both the **UploadedRpts** and **ErrRpts** folder contained within your personal folder. A successful attempt will result in the user's stations template being moved from the Stations folder to the user's **UploadedRpts** folder on the ftp. The file will be renamed with a date and time stamp along with the original name of the file. You may need to refresh your view to see the updates on the ftp.
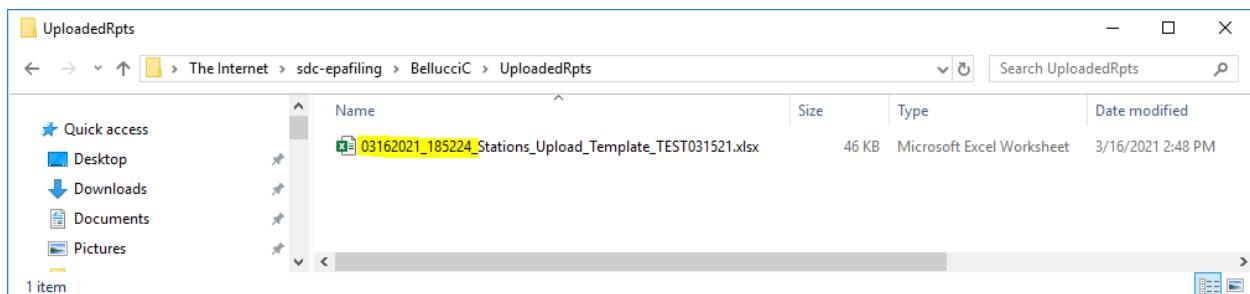

Figure 10. A successful attempt at uploading data will result in the original file being moved to the UploadedRpts folder and renamed with a date and time stamp before the original file name.

Rows of data that failed database constraint checks (QA checks) will be listed in an error file found in the **ErrRpts** folder. For more information on error report, see section 6.0 for more information on error reports.

6.0 Error Reports

An error report is automatically produced every time the system attempts to upload new data to the database. The error report produced is a tab-delimited file that can be open in any text editor or excel.

N.B. You need to make a copy of the error report onto a local or network drive

If all rows of data submitted by users are successfully inserted into the database the error report will produce a file that notes:

'All rows successfully inserted'

If there were rows of data from the upload template that were not successful inserted, each unsuccessful row will be identified in the error report with an error message. The first three columns of data indicate

1.) **The original file name (Example- '*Stations_TEST_data.xlsx*')**
2.) **The row from the original file that was not inserted**
3.) **An error message indicating why it was not inserted into the database**

The columns following the first three columns will be a copy of the data in the noted row from the upload template. Users can check this information to determine the error and resubmit in a new template file.

Figure 11. Example of an error report opened in Excel. The first three columns indicate the file name, row from the original file that contained an error and an error message from the database.

_____

The database produces a variety of errors to maintain data integrity. Table 2 provides a list of some common errors that users may come across and a more detailed description of the error that may help to fix issues with the data leading to a successful insert of that row.

Table 2. Common database error messages found in the error reports, a more detailed description of the error and how to resolve the issue.

Database Error Message	Description of Error	How to Resolve
1062 (23000): Duplicate entry '41.4096--73.2855' for key 'stations.PRIMARY'	A duplicate entry error indicates that the fields that make that row of data unique are already in the database. A duplicate entry of data is not allowed.	Check to make sure the entry is not a duplicate of data that is already in the database.
3819 (HY000): Check constraint 'xlong_outOfBounds' is violated.	The longitude value is not in or around Connecticut.	Check the longitude value.
3819 (HY000): Check constraint 'ylat_outOfBounds' is violated.	The latitude value is not in or around Connecticut.	Check the latitude value.
1216 (23000): Cannot add or update a child row: a foreign key constraint fails	One or more of the 'pick list' fields in the template was entered incorrectly OR a required field is missing.	Check 'pick list' fields to make sure values entered are listed in the 'pick list'. Check the FieldDescription worksheet in the template to identify what fields are required and make sure those fields are entered in the template.