User Manual

December 12th, 2024

Table of Contents

I. Introduction

MSConnect is a Web-based, flexible, open-source platform for automated high-throughput MS-based omics. It streamlines the entire process from data collection, data management, data processing and to data visualization by integrating and supporting various third-party tools, allowing for a fully autonomous. The platform is built with Python, Django, JavaScript, and HTML and works with Raw File Uploader and Processor.

What It Does:

MSConnect is a comprehensive platform that automates the following tasks:

- **Data Management**: Streamlines the uploading and organization of raw MS files.
- **Data Backup**: Provides secure storage throughout the processing pipeline.
- **Data Processing**: Interfaces with third-party tools to analyze MS raw data.
- **Result Interpretation**: Generates tables and figures for easy interpretation and presentation of results.

Key Features

- **Vendor-Independent**: Compatible with diverse MS applications, ensuring flexibility across platforms.
- **Automation**: Offers fully autonomous workflows, minimizing manual intervention.
- **Third-Party Integration**: Seamlessly connects with external software for advanced analysis.
- **Visualization Tools**: Provides clear and actionable data representations for streamlined interpretation.
- **Generalized Framework**: Adaptable to various omics studies, expanding its applicability.

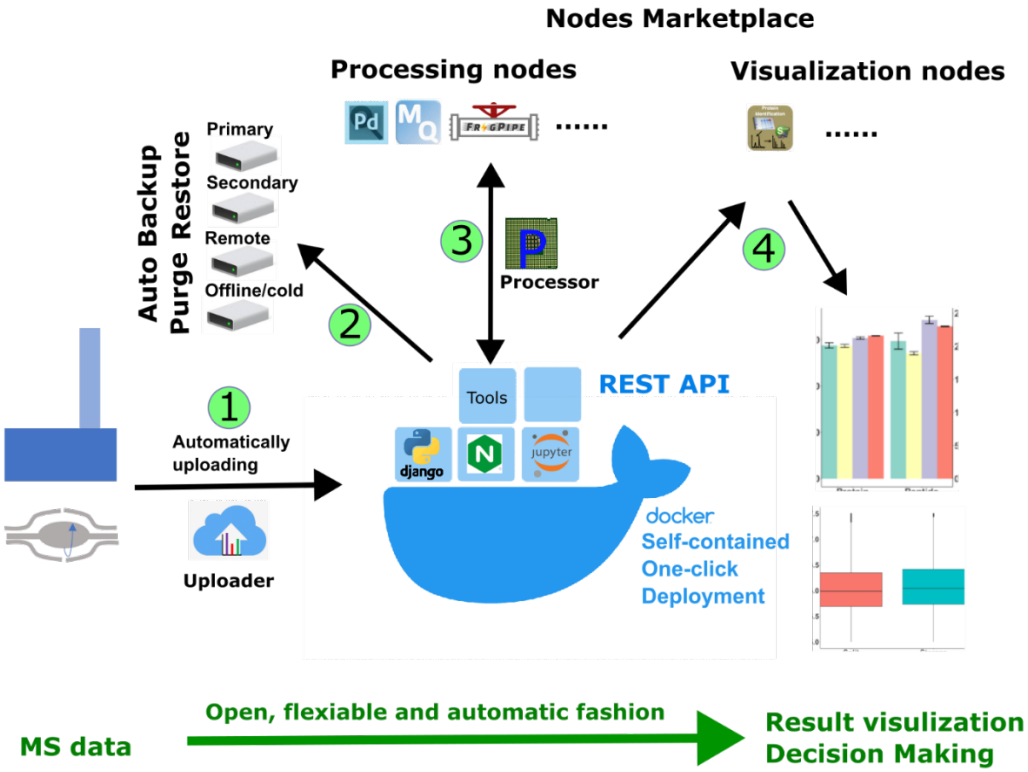
What It Is Not

- MSConnect is not a standalone data analysis tool. It depends on third-party software for data processing/analysis
- MSConnect does not come with any data processor or analysis software/license, it provides a wrapper/interface to interact with them.
- It is not limited to any specific MS application; instead, it provides a generalized framework adaptable to various omics studies.
- MSConnect is not a desktop application; it is web-based and requires a server environment for deployment.

The following diagram illustrates the key steps in the data management platform.

1. Mass spec data is automatically transferred from individual control PCs to the server through the Raw File Uploader. The upload process is vendor independent and can work with any MS system that uses files or folders for storage, as the Raw File Uploader performs the upload based on OS-level file changes, rather than vendor-specific scripts. For more information, see the Raw File Uploader.
2. Once the server receives the raw file, it creates a SampleRecord with all the metadata and converts the data using a Docker command (the default is https://github.com/phnmnl/container-pwiz). It also creates a cache file for display of the MS1 plot and saves the files to up to four storage devices based on user settings.
3. If a default data processing protocol is specified, a processing task will be automatically added to the Data Analysis Queue and processed by processors installed on data processing computers. The results will then be uploaded to the server once processing is complete.
4. Finally, users can view the processed data through visualization nodes or through the R or Python environment in Jupyter Notebook.

To ensure data redundancy, all raw and processed data is automatically backed up to different storage devices as specified during installation.



II. System Requirements

Hardware

The platform can be installed on any hardware that supports Docker, as we have tested it on a variety of systems including a Raspberry Pi 4 with 4GB of RAM, a desktop PC running Windows 10, an Ubuntu server running version 22.04, an ESXI virtual machine, and a cloud deployment. It can run on Linux, Unix, and Window OS. We recommend having at least 8 GB of ram and 120 GB of storage.

Software

- Raw File Uploader (RTKlab-BYU)
- Proteomic Data Processor (RTKlab-BYU)
- Licensed third-Party proteomic identification software, such as DIA NN or Fragpipe

Internet/Network

MSConnect server and its subprogram, Raw File Uploader and Proteomics Data Processor can all be installed on the same PC. In this setup, no internet or network connection is required. However, it is recommended to install the MSConnect server, Raw File Uploader, and Proteomics Data Processor on separated PCs to reduce the load of individual computers, especially the MS controlling computer. For this setup, an internet or network connection is then necessary.

Supported Mass Spectrometers

This platform supports virtually any file or folder-based MS system. The Raw File Uploader performs uploads based on OS-level file changes, rather than vendor-specific scripts, making it compatible with any MS system that uses files or folders for storage.

The conversion and extraction process use a 3rd party Docker container, with the default tool (https://github.com/phnmnl/container-pwiz) already supporting many platforms. Any other Docker image can be used, and the conversion and extraction process is not strictly necessary, providing more information and flexibility.

III. Different Type of User

Admin

The admin installs and maintains MSConnect, Raw File Uploader, and Proteomics Data Processors. The admin has comprehensive access to the website for advanced settings and plays a critical role in supporting end users and troubleshooting. Examples of admin include MS facility manager, lab manager, and lab IT

End user

The end user mainly interacts with Raw File Uploader and the MSConnect website. The end user has access to manage records uploaded under his/her account, and the automatic QC preset selection. Examples of End user include Research scientist and Research Assistant

Developer

MSConnect is open source. Program developers can modify it to meet specific needs that are not currently implemented.

IV. MSConnect Installation

Three major programs are required for the best utilization of MSConnect. MSConnect is the data manager platform, Raw File Uploader is used to transfer MS files to the MSConnect server, and Proteomic Data Processor allows for automatic analysis using the different vendors' identification software and provides a quick view of the quality of individual MS runs. This section provides details outlining how to set up each program.

Setting up MSConnect server

The platform is installed as Docker containers, so you need to have Docker software installed first. The installation takes 5 steps and should take less than 5 minutes. User can also reference the step-by-step screenshots guide.

1. Install Docker Compose

Follow the installation guide from Dockerdoc (https://docs.docker.com/compose/install/). For Windows systems, Docker Desktop must be manually started. You can do this by clicking the Docker Desktop icon, which allows the application to run in the background. In older versions of Windows, if Docker prompts you to set up Windows Subsystem for Linux 2 (WSL 2), you only need to follow steps 4 and 5, not step 6. Once these steps are completed, a system reboot is necessary for the changes to take effect.

2. Download MSConnect

Download MSConnect by clicking on this link, unzip the folder, and put in the desired directory.

3. Configuring .django_secrets.env

Rename .django_secrets.env_example to .django_secrets.env (don't forget the "." at the beginning of the file name, only delete the "_example" part). You can configure certain parameters used within the Django app, such as time zone, debug mode, email server, and Jupyter password. These parameters can be adjusted based on your specific needs. If you're not sure what to set, you can leave them as their default values or configure them as their names suggest. It's important to ensure that any changes you make align with your system settings and operational requirements.

4. Configure docker-compose.yml for storage location

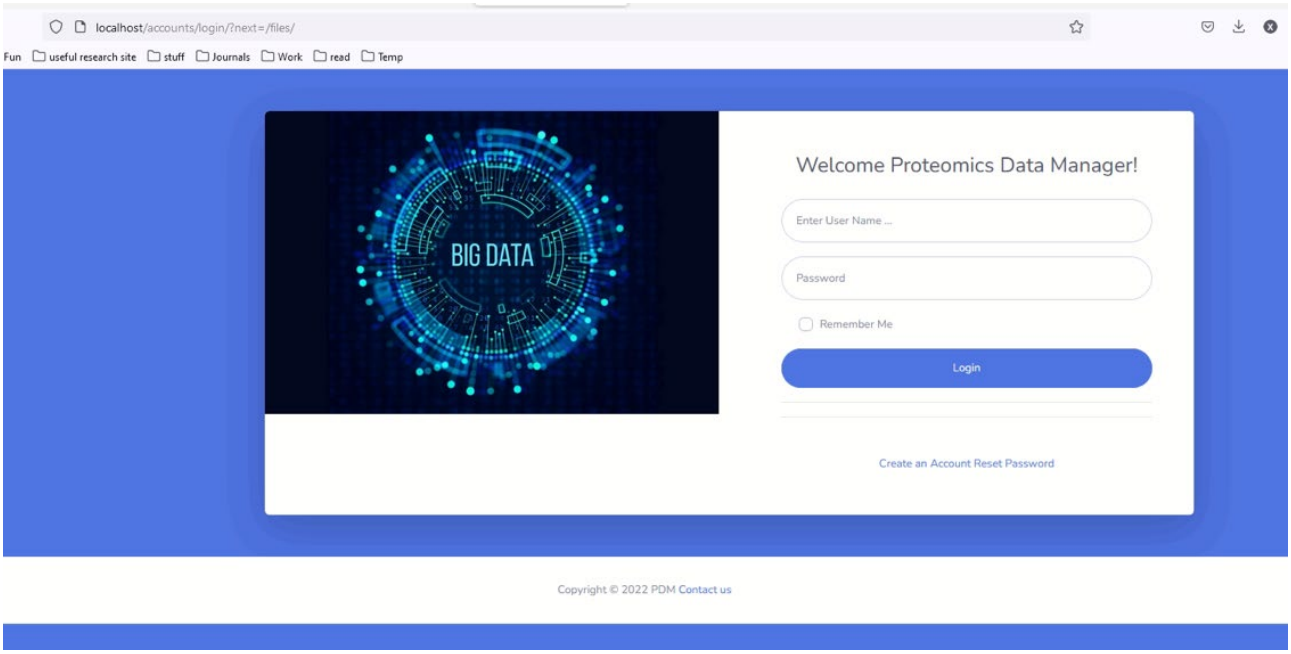
Configuring the storage locations for your data, including raw files and processed data, involves setting up to four different locations or drives. The primary storage is a required configuration, while the other three - secondary storage, remote storage, and offline storage - are optional but highly recommended to ensure redundancy. If the specified folders do not exist, the system may return an error. Thus, to ensure compatibility, all four drives are set to the 'C' drive in a Windows system by default. However, you can change any of these storage locations by modifying the corresponding 'source' field in the 'bind' section (only change the 'source' field; do not alter the 'target' field). (See screenshot)

The format for specifying storage locations differs based on your operating system. For Windows hosts, the format is /d/docker/primary_storage, which corresponds to the location d:/docker/primary_raw. Linux, on the other hand, uses its native format, /user/xxx/documents/primary_raw. You can configure up to four storage locations (primary_storage, secondary_storage, remote_storage, offline_storage) for use in the app. If you're not using all of them, you can either comment out the unused ones using "#" or point them to an empty folder (recommended for easy future expansion). The naming conventions for storage locations, such as "remote_storage," do not need to be taken literally. For instance, "remote_storage" does not necessarily have to point to a remote folder. These are default names and can point to any location based on your preference and requirements.

5. Run docker-compose up

After you have configured the necessary files, you can start the app by running the command "docker compose up" in the directory of the extracted folder (usually named MSConnect-master). You can do this in the Command Prompt for Windows (Guide) or the Terminal for Mac or Linux (Guide). If you encounter a "file exists" error, simply retrying the command usually resolves the issue. Alternatively, if you have installed Visual Studio Code or another integrated development environment (IDE) with the appropriate extensions, you can right-click the docker-compose.yml file and select "Compose Up". If you encounter any errors, refer to the troubleshooting page for potential solutions and advice, or contact us at Github page.

The website home page after proper installation looks like the Figure below, and now your server should be functional.

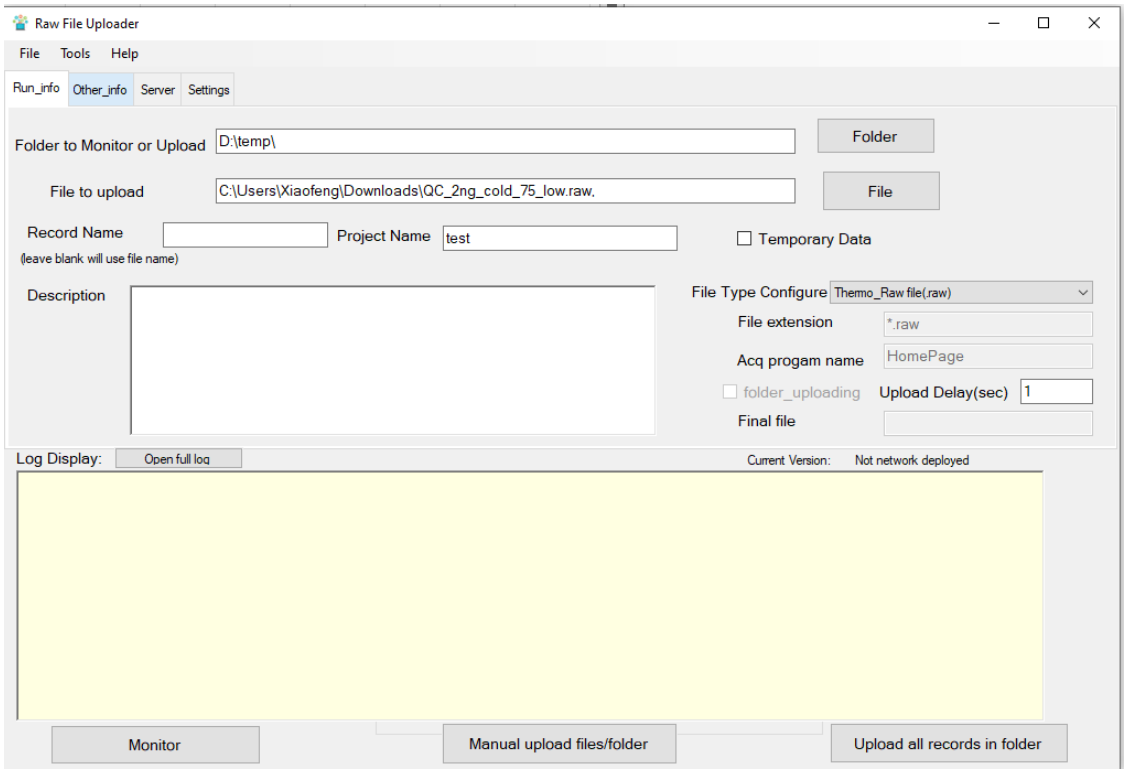


MSConnect website Login page

Setting up Raw File Uploader

Raw file uploader is an upload tool to be used with MSConnect for uploading raw files to the database with metadata about the analysis. The program is built on C# with .net 4.5.1 (same framework as Thermo Orbitrap Tribride) for the Windows platform. To setup the Raw File Uploader, follow the instructions below:

1. Download the Raw Files Uploader
   You can download the .zip file containing the Raw File Uploader from GitHub or click Download for it to download directly from your browser. With the .zip folder, follow these steps
2. Open the .zip folder
3. Click and run the setup application
4. If you are restricted by security, click more info, then click "Run Anyways"
5. Once complete, the Raw File Uploader is installed.



Raw File Uploader-Run_info tab

Setting up Proteomic Data Processor

The Proteomics Data Processor is a Windows-based application designed to integrate with MSConnect and automate the processing of mass spectrometry data using third party identification software, such as FragPipe, DIA-NN, and ProteomeDiscoverer. It provides users with real time feedback on the quality of MS analysis, such as the number of protein and peptide identification to aid decision making process. This tool works in tandem with MSConnect to process the MS-based proteomics data. The program is built on C# with .net 4.5.1 (same framework as Thermo Orbitrap Tribride) for Windows platform.

The Key Features of Proteomics Data Processor include:

- Seamless communication with MSConnect.
- Automation of data processing through desktop tools.
- Designed for Windows systems with .NET Framework 4.5.1.
- Command-line configuration for admins and user-friendly GUI for researchers.

1. High-level overview

The processing workflow connects MSConnect with third-party tools like FragPipe and DIANN for proteomics data analysis. Three components listed below are critical for the workflow :

- **Wrappers:** The tool installed to MSConnect server that bridges between MSConnect third-party applications, and processors/worker.
- **Processors/worker**: It is the Proteomics Data Processor that downloaded and installed to the computer and is used to execute workflows configured in MSConnect server.
- **Processing Workflows**: Define the structure and logic for MS file data processing using the processor/worker.

The Proteomics Data Processor can be installed on multiple PC to create more node for Parallel Processing. This section provides a comprehensive guide for setting up the processing workflow after the MSConnect server has been initialized. Note that the Proteomics Data Processor is also referred to as Processing application, or Processing worker.

2. Installation instructions

The installation process involves downloading and configuring wrappers, setting up workflows for processors, and preparing the Proteomics Data Processor application.

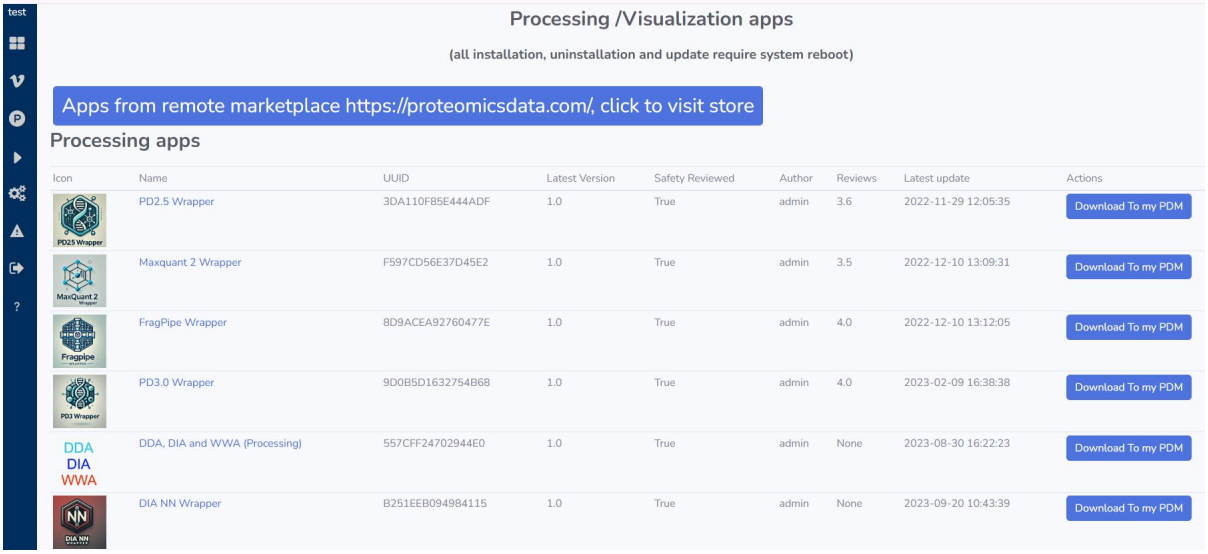
a) System Requirements

Before proceeding with installation, ensure the following requirements are met:

1) **Hardware:**
   - Windows OS (Windows 10 or higher).
   - Administrator privileges on the machine.
2) **Software:**
   - .NET Framework 4.5.1 ([Download here](https://dotnet.microsoft.com/download)).
   - Processing tools with valid licenses:
     - [FragPipe](https://fragpipe.nesvilab.org/)
     - [DIA-NN](https://github.com/vdemichev/DiaNN)
     - [ProteomeDiscoverer](https://www.thermofisher.com/).
3) **Network:**
   - Ensure internet connectivity for MSConnect communication.

b) configuring wrappers to MSConnect server

i) Login to MSConnect as admin:
ii) Download Wrappers:
   (1) Navigate to the Processing tab.
   (2) Click "Get More" at the top of the page to access available wrappers.
   (3) In *Apps from the remote marketplace* section click "Download to my PDM" under the action column of the desired wrapper (e.g., FragPipe, DIANN, or PD).
   (4) Scroll down to the bottom of the page and click Install on the app you desire to install in 'Currently Downloaded Apps in Local Proteomics Data Manager'.
   (5) Once the app is installed, click enable under the same section where you
iii) Restart the Server:
   (1) Go to the Advanced Settings page under the Settings tab.
   (2) Press ‘Restart Data System’ to initialize changes.
   (3) Wait a few minutes for the restart to complete.
iv) After the restart of the server, the Processing Wrappers should now appear in the Processing tab at the top.



MSConnect Website-Remote Markerplace for Wrapper downloading

c) Create Processing workflow

The Processing workflow is called Automatic QC setting preset on the MSConnect Settings page. The selected processing app/wrapper and preset will be uploaded to the processing queue along with the MS files via raw file uploader.

The installation of a wrapper creates a ProcessingApp object associated with the wrapper under the admin page. It is preloaded with a couple presets that users can use as template to modify their search parameter, such as FASTA and acquisition mode. The following instructions show how to create a different preset that uses different search parameters, such as FASTA from an existing search result from the third-party program such as DIA NN or Fragpipe

1) Go to the **Admin page** under the Help tab.
2) Go to the '**Processing Apps**' section under the FILE_MANAGER
3) Select the desired **'ProcessingApp Object'**. The ProcessingApp object associated with the specific processing wrapper will be created on the admin page after the installation of wrapper on MSConnect website. By default, each of the apps are called ProcessingApp object (#). You can find the name of processing app/wrapping once clicking into it. For example, on the demo site:
    - ProcessingApp object (2) : PD 3 Processor
    - ProcessingApp object (3) : FragPipe Processor
    - ProcessingApp object (4) : DIA NN Processor

4) **Edit Preset**. There are 8 preset slots and 2 user preset slots. The preset slots will be replaced if the wrapper is updated to a newer version, while the user preset will not. Each preset is a zip folder that contains all input files defined parameter for a specific processing app/wrapper. The content required in the zip folder is different between wrappers.
   - **Create a new preset:** upload the zip folder then click save at the bottom of the page.
   - **Delete an old preset**: check the clear box at the right of preset then click save at the bottom of the page.
   - **Edit an existing preset**: replace the zip folder by uploading a new zip folder next to change then click save at the bottom of the page.
   - **Build a zip folder:** the input files in the folder define the search parameters and is generally one of the output files resulting from the search on the third-party identification software.
     i) Run the search on local identification software.
     ii) Export required output files from identification software as input files sources
     iii) Rename the output files to match exactly the name given in the preset template, also listed in the table below.
     iv) Put all the files into a zip folder. The folder can be named any way for easier recognition.

Identification Software	Input files name in .zip folder	Input files sources
PD 2.5/PD3	input_file_1.pdProcessingWF	pdProcessingWF
	input_file_2.pdConsensusWF	pdConsensusWF
FragPipe	input_file_1.workflow	fragpipe.workflow
	parameters.json	Manually change the data_type to DDA, DDA+, or DIA
DIA NN	input_file_1.txt	From 'report.log.txt', Copy the parameter (and starts after the input file name and end at the end of that section of log) and save it as a new . txt file.
Note: Make sure the path and files for FSATA and Speclib on computer that house processor worker is the same as in the input files, and is identical across all computer that execute parallel processing to ensure program functioning correctly and consistency results.		

d) Preparing the Proteomics Data Processor application

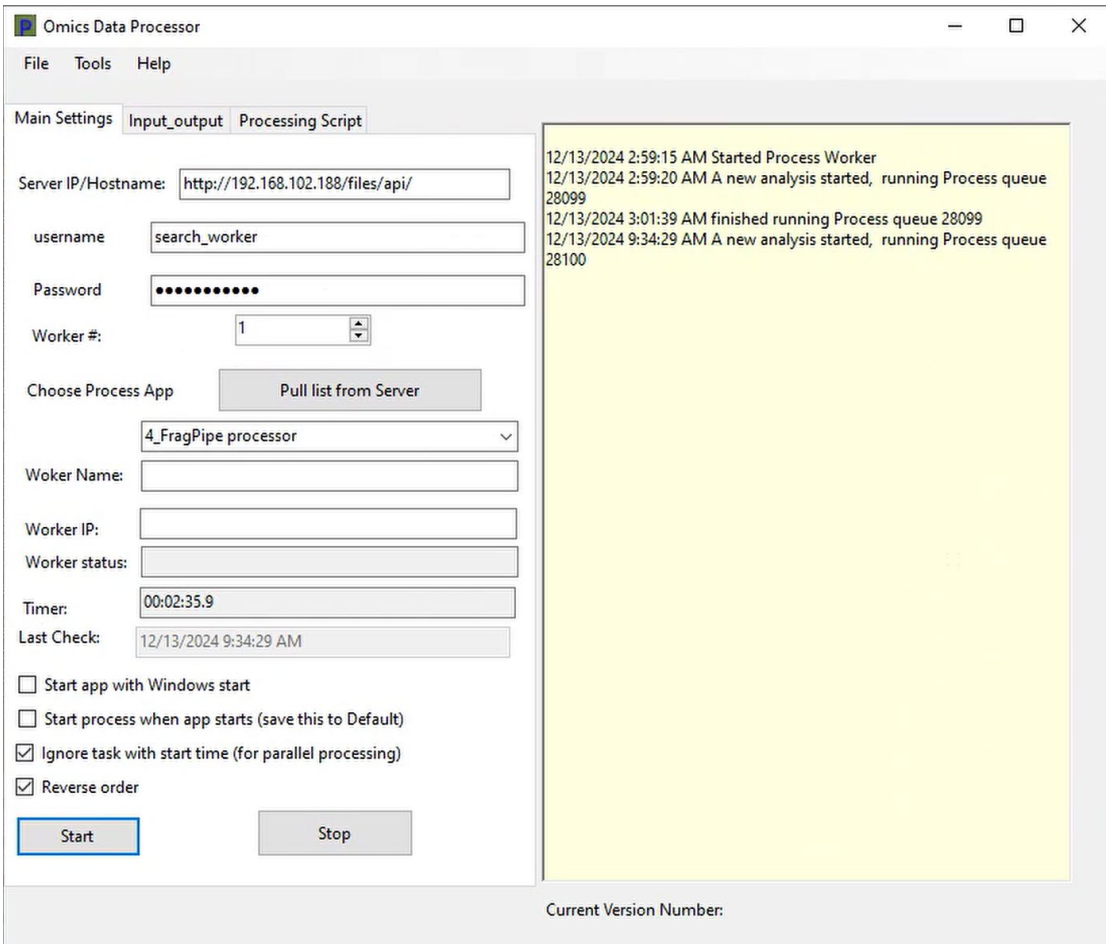
The Proteomics Data Processor is the application installed on the computer that houses the third-party identification software like DIA NN, Fragpipe, or Proteome Discover. It is also referred to as the worker in this documentation. The worker downloads the MS files and the processing preset from MSConnect server and connect it to the identification software via command line. It can be installed on multiple computers for Parallel Processing and get through the queue faster.

1. **Download** the application from GitHub
    a. Click on the most recent release on the right panel
    b. Download the Protemoics _Data_Processor_1_7.zip
2. **Launch** the application (.exe) from the zip folder after extraction and create a shortcut on the desktop.
3. **Edit application setting**. The application automatically loaded `default.xml` upon starting, this setting file contains parameters that connects the application to the MSConnect server and to the local identification software. The setting files application is specific to individual wrapper. In this section, the setting of the three tabs on the application are explained. At the bottom of this section, you can find the sample setting files for DIA NN, Fragpipe, and PD wrapper.

a) Main Setting

- Connect the application to MSConnect server by entering the following information
    i. Server IP/Hostname.
    ii. Login credentials
    (user: `search_worker` , password: 'searchadmin')
    iii. Computer name and IP address.
    iv. Assign a worker number for identification
- Configure Application Settings
    v. Click pull list from server
    vi. Select the process app from the dropdown menu. Only wrappers that are installed will be shown in the dropdown menu.
- Select the checkbox you want to implement for this specific worker
    vii. Start app with Windows start.
    viii. Start process when app starts.
    ix. Ignore tasks with start time (for parallel processing).
    x. Reverse order (queues data to process the newest runs first

- Provide worker status.
    xi. Time stamp when an analysis queue started and finished
    xii. The queue number for the analysis being processed
    xiii. The total time spent for the analysis being processed
    xiv. Worker status
    xv. The last time that the application being checked by the server.



Proteomic Data Processor-Main Settings tab

b) Input_output

- Create Temporary Folders that store the MS file to be processed and the output files generated by the identification software. (e.g., `D:\QC_DIA`). Files in this folder will be cleared once the analysis finished and all the output files are uploaded to MSConnect (shows in Processing page).
- In each of the Output slot (1 -6), indicates the output files path and name under the temporary folder folder you would like to upload to MSConnect. The file name must match exactly for successful uploads. For example, `report.pr_matrix.tsv`, `report.pg_matrix.tsv`, `report_stats.ts`,and `report_log.txt` are uploaded as output file 1 to 4 in the example screenshot for DIA NN wrapper.

c) Processing Script.

The processing script is the command line that sends the search request and parameter to the identification software. The script is varied between wrappers. Below are the examples script for different wrappers. The most critical variation from user to use is the path to the execution file of the identification software of the choice.

- **Fragpipe:**
  /c C:\Fragpipe_22\fragpipe\bin\fragpipe.bat --headless --workflow &&input_1&& --manifest &&input_2&& --workdir D:\QC_fragpipe\ --config-tools-folder C:\Fragpipe_22\fragpipe\tools\
- **DIA NN**
  /c C:\DIA-NN\1.9\DiaNN.exe --out D:\QC_DIA\report.tsv &&loop&& --f &&raw_file_name&& &&loop&&
- **Proteome Discover**
  /c DiscovererDaemon.exe -c custom &&loop&& -a custom &&raw_file_name&& &&loop&& -r &&output&&.msf -b -e custom ANY &&input_1&&;&&input_2&&

d) Setting files example. Follow the instructions to tailor the setting to your need and save it as `default.xml` in the file location of the application.

- Default_DIANN.xml
- Default_Fragpipe.xml
- Default_PD.xml

4. **Setup the application for different wrappers.** There are two approaches to setup the application for different wrappers.
    a) **Duplicate the zip folder (recommended).** This approach is more straight forward to the end users. Since you can rename the shortcut according to the specific wrapper, and the users also don't have to worry about choosing correct setting. It also allows parallel processing for different wrapper on the same PC.
        1) Duplicate the zip folder
        2) Rename the folder and .exe for easier recognition.
        3) Edit the default setting files to meet your need following instruction above.
        4) Save the new setting files as `default.xml` in the file location of the application
        5) Create a shortcut on the Desktop
    b) **Load wrapper specific setting file** after starting the application. This requires more communication between admin and end user.
        1) Create setting file for different wrappers.
        2) Save it under the file location of the application with clear naming structure to differentiate between wrappers.
        3) Every time when opening the application, click File/Loading settings from, then select the desire setting files.

5. **Use the application for parallel processing.** The application can be installed on to different PC or be opened multiple times on the same PC for parallel processing. To installed on different PC, follow the steps above. To execute parallel processing on the same PC, simply double click on the application again, and it will start a new window. The following two notes are critical for successful parallel processing.
    a) Make sure the path and files for FSATA and Speclib are identical as in the input files of the preset across all computers that house processor application/workers .
    b) The worker number is different between all processing application/worker.

6. **Common sources of error.**
    - **Failed to verify workflow post set up.** Always test the worker and preset after setting up or making edition.
    - **naming conventions and folder structures** doesn't match the preset on MSConnect.
    - **The application/worker was not started.** This happens a lot when starting new set of data collection on MS after a period of idle, or when needing to switch to a different wrapper. It also happens when the PC is restarted, but the start app/process with Windows start box isn't checked. The best solution is to always leaving the application open and idle on your machines to ensure that it will pick up the next available task on the queue.
    - **The licenses of identification program is expired or the version doesn't match the processing script.**

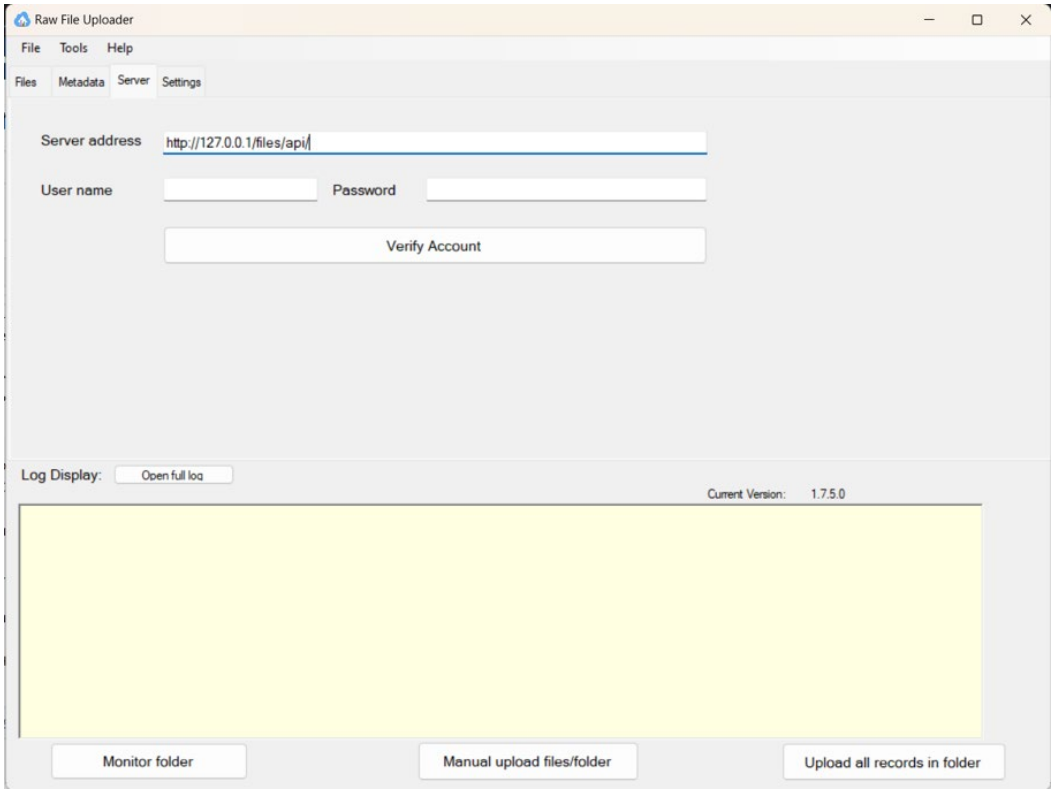
V. MSConnect Usage

Data Uploading -Raw File Uploader

1. Connecting to MSConnect (Server Tab)

Before you can upload files to MSConnect you need to set up a pathway for the Raw file uploader to reach MSConnect. For this, you'll need the server IP address for MSConnect and your Username and Password.

1. IP address for MSConnect
    a. If MSConnect is running on your current device then you can use one of these Addresses:
        - http://127.0.0.1/files/api/
        - http://localhoast/files/api/
    b. If MSConnect is running on a different device replace "IP_Address_Here" with the IP Address that hosts MSConnect : http://IP_Address_Here/files/api/
2. Open the Raw File Uploader, enter the MSConnect IP address, put in your Username and Password, and click "Verify Account".
3. A success message will pop up which indicates a successful connection. Now you should be able to upload files to MSConnect.



Raw File Uploader-Server tab

2. Setup Configuration (File Tab)

Before you upload it is good practice to record important information for your colleagues and your future self. These settings also tell the Uploader and MSConnect how to process this information.

- ***Project name****:* This is where you write the name of the project that you are working on. It will create a folder in the storage.
- ***Batch Name****:* This is a subclass of the project, naming a Batch helps identify what is being tested. It is good practice to write the date or some other type of identifier to help you distinguish batches.
- ***Enable batch name:*** Checking this box will create a separate folder to store this batch in
- ***Assign to other user*****:** After selecting "Pull user List" this will allow you to assign the file uploaded to a specific user
- ***Description:*** Use this to write important information about your batch such as, the description of the project, the specifics of the batch, and the date on which these things were performed. *The more information you record the easier it is for others to understand your data*
- ***Temporary Data****:* Enabling this will preset your data to delete after 3 months
- ***File Type Configuration****:* This is a setting that tells the uploader what type of data you are uploading.
- ***Upload Delay (Sec):*** This setting provides buffer time between uploads.

3. Three uploading modes(File Tab)

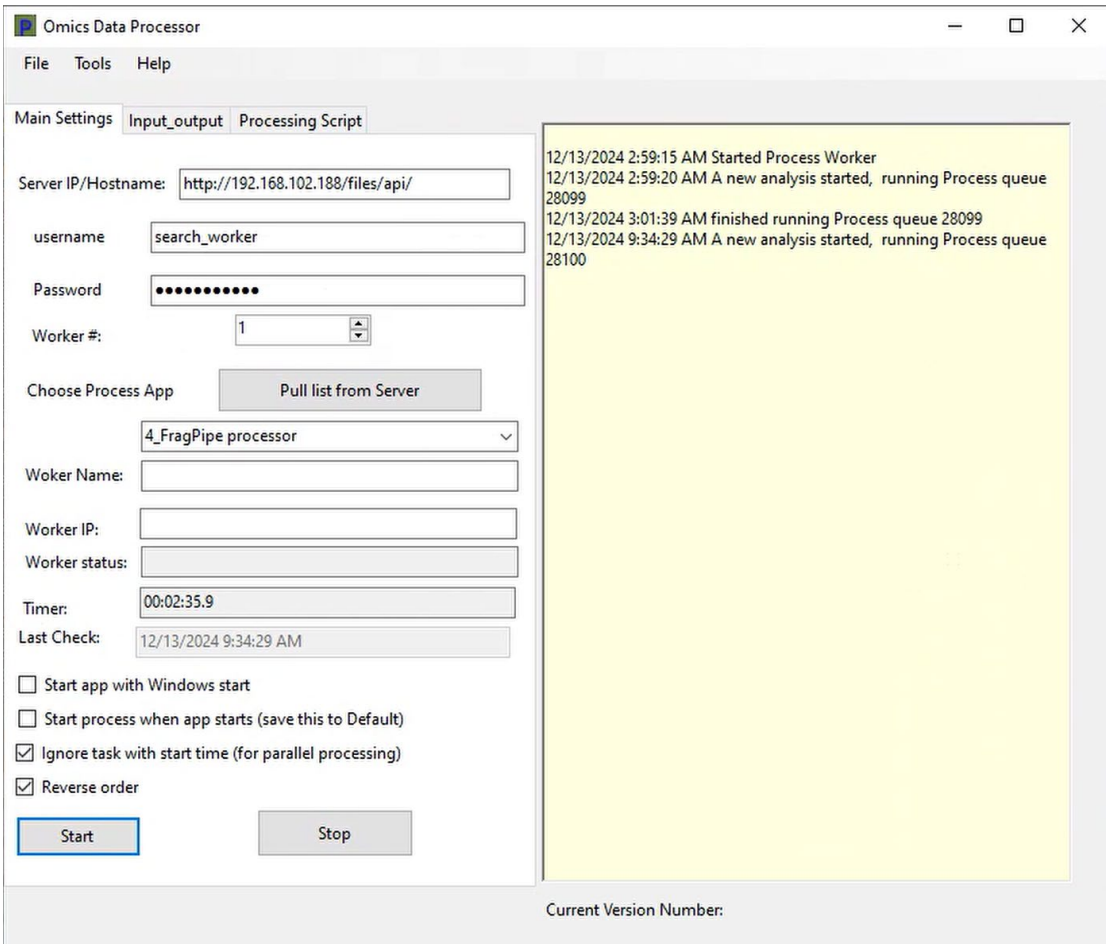
1. **Monitor the entire folder during MS data acquisition**
   Browse to the desired folder under 'Folder to Monitor or Upload'. Then, click the 'Monitor Folder' icon. The icon will turn green indicating it is being monitored, and the MS file will be uploaded to MSConnect automatically once the acquisition is completed.
2. **Manual upload file/folder**
   Browse or type a filename (or files separated by '.'). then click 'manual upload files/folder'. It supports multiple files selection.
3. **Upload all record in folder**
   Browse the desired folder under 'Folder to Monitor or Upload'. Then click 'manual upload files/folder'.

Data Processing- Proteomics Data Processor (Worker)

1. Double-click the desktop executable to open the application. The admin should have already set up the processor so it will contact the MSConnect server and the chosen Process App automatically upon opening.

Proteomics Data Processor icon on desktop

2. Confirm the Process App is the desired one.
3. Confirm that the 'ignore task with start time' and 'reverse order' boxes are checked
4. Click Start to start the process worker.
5. Monitor MSConnect queues and troubleshoot as needed.
6. It is suggested to always leave the application open and idle on your machines to ensure that it will pick up the next available task on the queue.



Proteomic Data Processor-Main Settings tab

Data Management-Website

1. Accessing the website

a) Address

There are two ways to access the MSConnect website. It can be accessed on the device that it was installed or it can be accessed over the internet.

- **Accessing the application locally** After installation, it's a good idea to first access the application locally to ensure proper installation. Users can access locally (on the same installed system) by typing "localhost" into your browser's address bar. A login screen should appear, then you can use the below default admin account or create a new account to access the application.
- **Accessing the application through the network** Instead of putting the localhost in the browser's address bar, you need to put the IP address of the server (where the application is installed). Find the IP address of a PC with these help links, Windows IP address, Linux IP address, Mac IP address

b) Gaining Access

By default, there are two accounts created preinstalled with the software.

- **Admin Account** - intended for application management. It's recommended to change the password for this account via the admin site once the application is deployed.

  Username: admin Password: proteomicsdatamanager

- **Generic Account** - standard user account, intended for database searching. Unless there are security concerns, this account can be left as is.

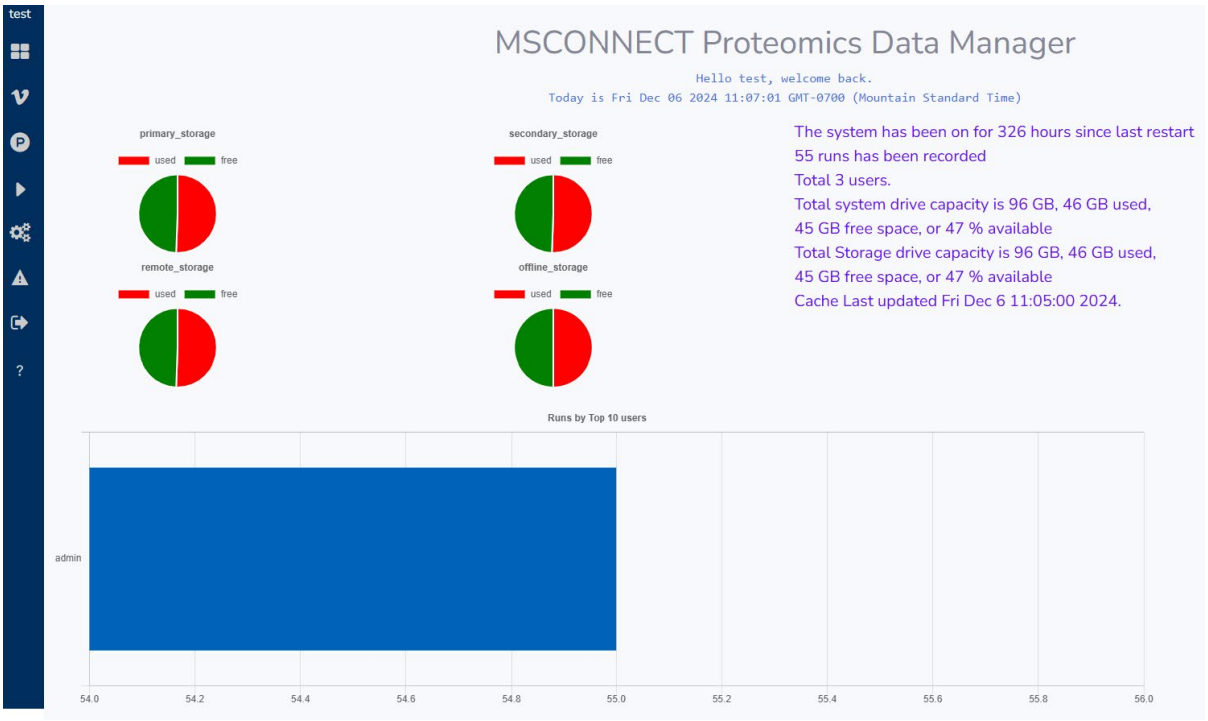
  Username: search_worker Password: searchadmin

- **Other Account** - User can sign up for a new account at the login page, and the admin can manager user access via the admin site

2. Webpage organization

MSConnect is organized into 8 different Tabs that each take the user to a different page that represents a different function. The main function of MSConnect is to take data and assist in the processing and visualization portions of the program. With each page comes a specific function.

A demo site was built to show the functionality of MSConnect. MS files previously collected in our lab were uploaded and processed for the demonstration. Note that the demo site has no processing functionality. Access the demo site at **http://proteomicsdata.com:8090** with username test, and password proteomics.



MSConnect website- Dashboard page

a) Dashboard Page- Display of Storage and user information

The Dashboard is the first thing a user is greeted with after logging in. The dashboard's main function is to provide an overview of MSConnect system usage. With this, page admins and users alike will be able to know if there is something wrong with MSConnect or if the storage is full.

Information displayed in the Dashboard page includes:

1. **In text:**
   - Hours since the last restart
   - Number of total runs recorded
   - Number of users
   - Each drive's capacity, usage, and free space
   - Last Update
2. **In graph:**
   - Drive Storage information
   - Numbers of Run by Top 10 users
   - Numbers of Run in the Last 30 days
   - Numbers of Run in the Last 12 months

b) Visualization Page- Hub for Mass Spec Visualization tools

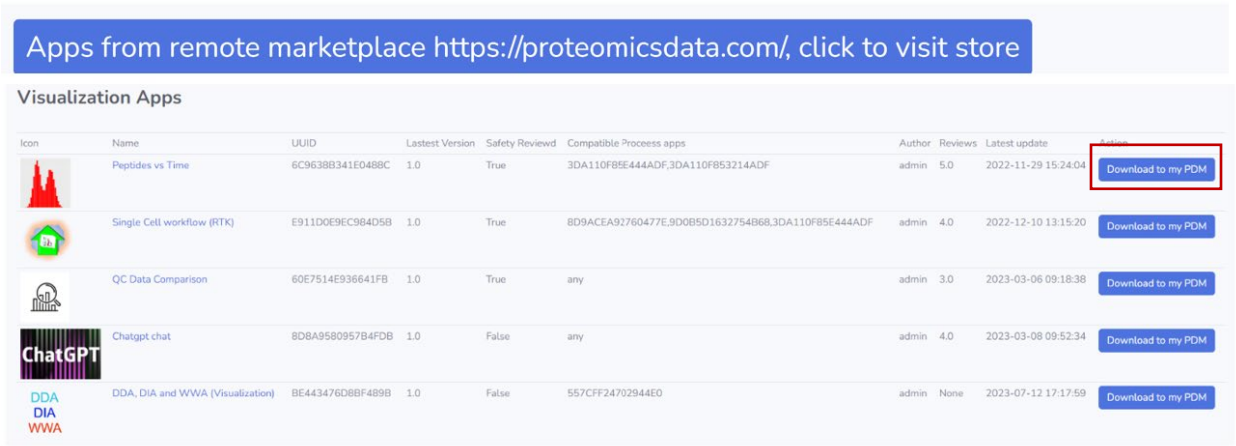
The visualization page displays a collection of tools that can be used to visualize data, such as comparison between different runs. When first installed, no Visualization tools will be present, they need to be installed by admin. To install a visualization tool, click the "Get More" button. This will take you to the App Center page. On the App Center, there will be multiple tools to choose from. Follow the steps below to install them.

(1) App Center (admin)

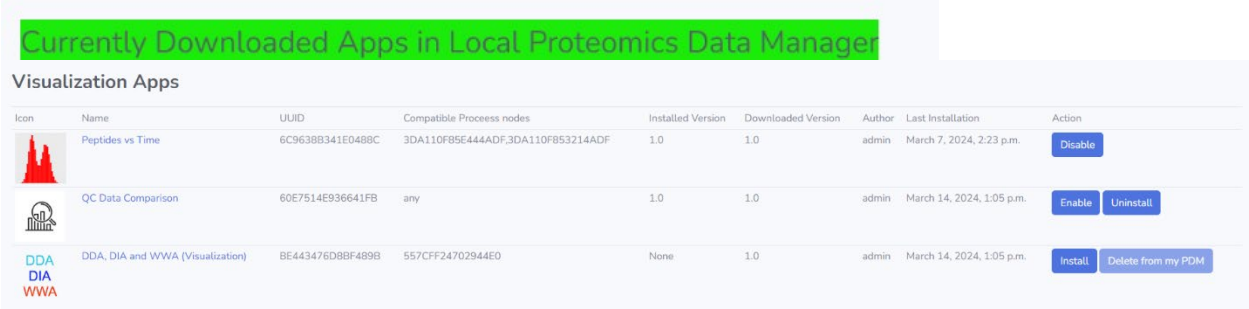
Before an App can be used it needs to be downloaded and installed to the MSConnect. The App Center has all the necessary tools to do so. For both Visualization and Processing tools, they can be downloaded, Installed, and enabled on this page. This page is a quick and easy way to disable a tool from MSConnect. An Admin login is required to access the App Center.

(a) How to Install an App

1. In *Apps from the remote marketplace* section click "Download to my PDM" under the action column of your desired app



2. Scroll down to the bottom of the page and click Install on the app you desire to install in 'Currently Downloaded Apps in Local Proteomics Data Manager'.



3. Once the app is installed, click enable under the same section where you click install.

Now the app will be present in the Visualization or Processing page. Both Visualization and Processing apps are installed the same way.

(b) Using Visualization apps

- QC data Comparison
- DDA, DIA and WWA

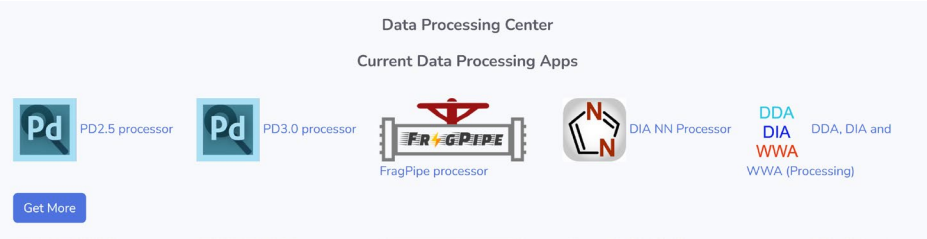
c) Processing Page - Hub for Mass Spec Processing programs

MSConnect allows for a quick examination of the quality of the MS run using different metrics such as the number of protein and peptide identification, PSM, and MS2 intensity via processing MS files. MS files can be queued for processing via two approaches: automatically via Raw Files Uploader or manually under Processing page. Aside from manually queuing up MS files processing, the Processing Page is also used to monitor the status of each processing worker and the MS files that are queued.

There are four sections on the Processing Page:

(1) Current Data Processing App (Wrappers)

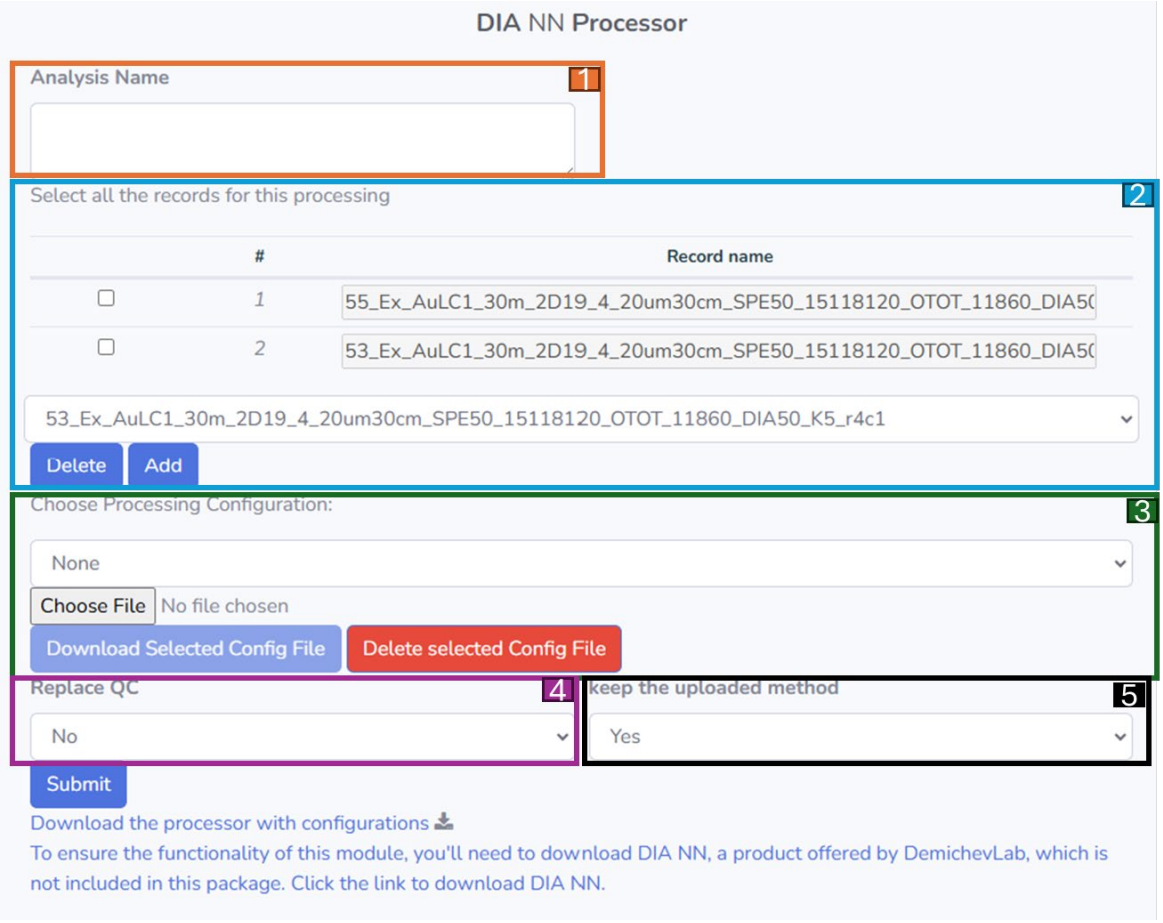
This section displays the Data Processing Apps (wrappers) that are currently installed on MSConnect as described in the App center section. The admin can install other processing apps (wrappers) that are provided in the App center by clicking Get More.



MSConnect website: Processing -Current Data Processing App.

To manually queue up MS files for processing, click on the desired processing app (wrapper) and following the instruction below:

1. Fill out the Analysis Name
2. Select records that have been uploaded to MSConnect using dropdown manual.
3. Choose the Processing Configuration file from the dropdown manual or manually upload one. See Creating an Automatic QC Setting Preset for the definition of a configuration file.
4. Decide whether to replace QC. When yes is selected, the processing result will replace the QC matrix under the Record page. If no is selected, the new processing result will only be display under Processing page.
5. Decide whether to keep the uploaded method. When yes is selected, the processing configuration file will be saved into the dropdown manual.
6. Submit



Steps to manually queue up for processing -DIA NN wrapper on Processing Page

(2) Current worker Status

A worker is defined as a Proteomics Data Processor installed on a processing computer which is associated with a specific processor apps/wrapper. Each worker is given a number, an IP, and a worker name. The worker's last job displays the last time that processor was used and the queue number that was processed.

(3) Search Queue

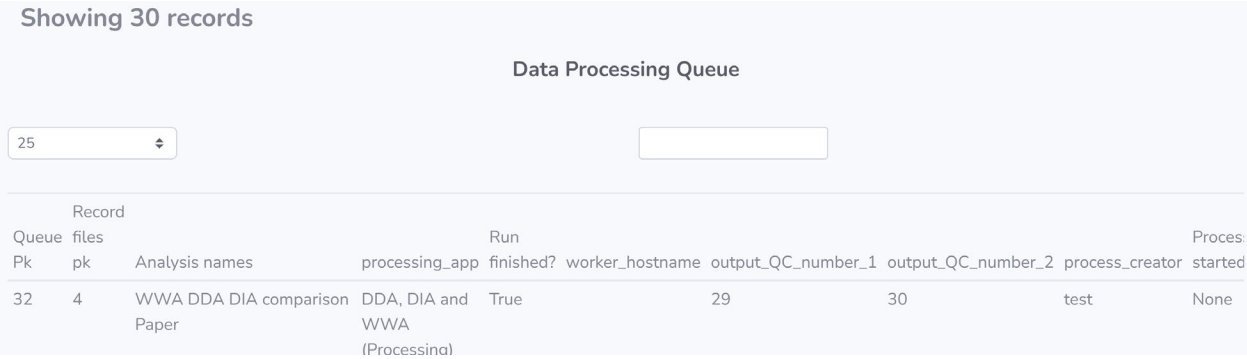
The last 100 MS files in the processing queue are listed in the Data Processing Queue by default and can be modified with search parameters to narrow it down to list specific data.

- **Queue Pk** - Searches for the Queue Pk number, which is the order that a MS record is queued up for processing. This can be used as a range.
- **Contain record** - Searches for the Record files pk number. The Record files pk number is the order when a MS file is uploaded to the MSConnect system through the raw file uploader. This can be used as a range.
- **Process Name** - Based on the Record Pk file name. It is shown as Analysis name in the Data Processing Queue.
- **Process Creator** - Once selected, only processes selected by this user will be displayed.
- **QC1 range** - Looks at "Output QC Number 1". QC1 matrix is defined in processing app/warper, and in most of cases linked to protein identification number. This can be used as a range.
- **Process apps** - Only displays the processed records that were created by the selected processing app/wrapper.
- **Finish time from and Finish time up to** - Used to set a time range that the process was run.
- **The search range can be put in three ways:**
    1. parameter to parameter. This is done by first inserting your smaller parameter followed by a hyphen and the bigger parameter. (ex: 300-500)
    2. from 0 to the parameter. This is done by inserting a hyphen followed by the parameter. (ex: -500)
    3. from the parameter to infinity. This is done by inserting the parameter. (ex: 500)

(4) Data Processing Queue

In addition to the matrix used in the search queue, the Data Processing Queue also contains below information:

- Run Finished? -If the record has been processed.
- Worker_hostname -indicated the worker processed the specific record
- The following matrix is defined under individual processing app/wrapper, and the output files can be downloaded (use the scroll bar at the bottom of the page)
    1. Output_QC_number_2
    2. Input_file1
    3. Input_file2
    4. Output_file_1
    5. Output_file_2
    6. Output_QC_number_3
    7. Output_QC_number_4
    8. Output_file_3
    9. Output_file_4
    10. Output_file_5
    11. Output_file_6



MSConnect Website: Processing-Data Processing Queue

d) Record Page-Provides Record Details

The **Records Page** is used to search and access specific record documents on the MSConnect Platform. Search parameters and the Secondary Search Function can be used together to pinpoint specific records. Record details can then be opened and read or the original file can be downloaded. At the bottom of the page, there is a table that displays the 100 latest runs.

(1) Record Table

The data table displays the last 100 runs recorded in the system. This table (orange box) will change according to the parameters set.

- **Pk** - When a run is recorded it is given a Pk or Record Number in the order it was recorded. Clicking on this number will redirect you to the Record Detail page.
- **Name** - Displays the name of the record
- **Protein** - Displays the QC number 1, pulled information from the data processor. The name of this column can be changed on the settings page.
- **Peptide** - Displays the QC number 2, pulled information from the data processor. The name of this column can be changed on the settings page.
- **PSM** - Displays the QC number 3, pulled information from the data processor. The name of this column can be changed on the settings page.
- **MS2** - Displays the QC number 4, pulled information from the data processor. The name of this column can be changed on the settings page.
- **Tool** - Displays the processing app/wrapper used to analyze the data.
- **File** - Displays the record's size of data file. Clicking the hyperlink downloads the file to your computer.
- **Acquisition** - Displays the time the MS acquisition was started. If no inquisition was made, displays when the record was first uploaded to the database.
- **Creator** - Displays the user that created the record.
- **Notes** - Displays any notes recorded on the record.

(a) Specific Search

Specify search parameters in green box can be used to modify the table at the bottom of the page to show more specific data. This can be used to isolate a specific record or to compare similar records.

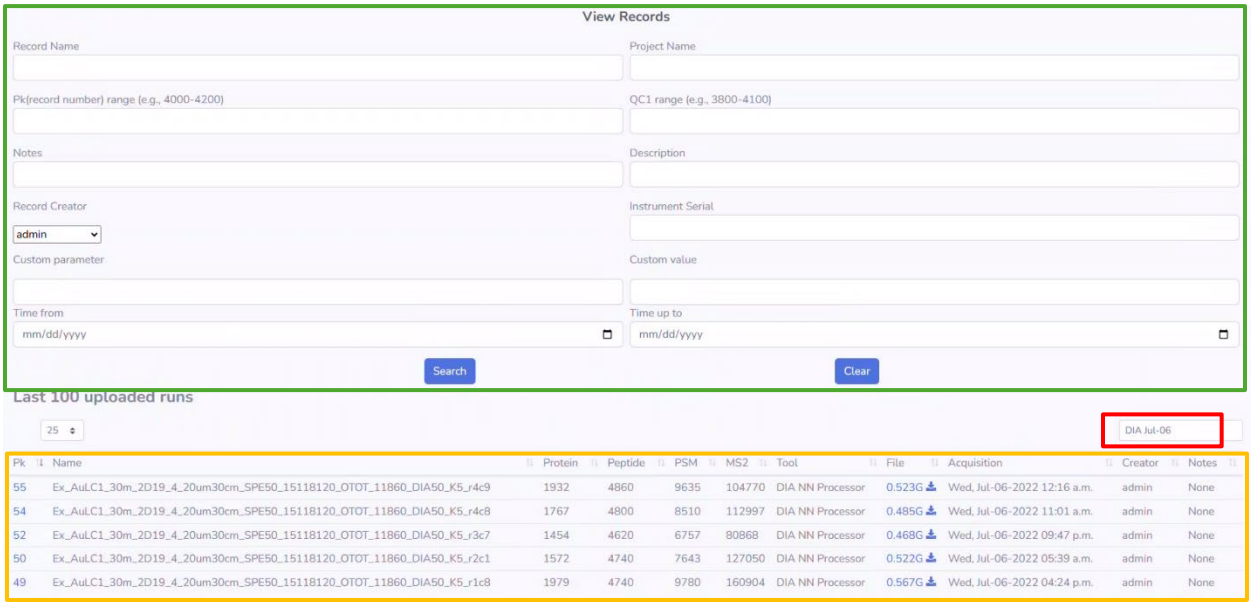
- **Record Name** - Search for keywords in the name of the record
- **Project Name** - Search for keywords within the display name of the project (This is imported from the Raw File Uploader)
- **Pk(Record Number)** - Searches for the record number. The record number is created when the record is uploaded according to what number record in the system it is. This can be used as a range.
- **QC1** - Searches through the **QC1 value** of each record. Can be set as a range. The QC1 by default settings represents the number of proteins in the record, this can be changed in the Settings Page.
- **Notes** - Searches for keywords in the recorded notes. Notes are added on the Record Detail Page

- **Description** - Searches keywords within the Description. (This is imported from the Raw File Uploader)
- **Record Creator** - Only displays records created by a specific user.
- **Instrument Serial** - Searches for a specific Serial ID given by the instrument that created the records data file.
- **Time from & Time up to** - Narrows display to only show records created between these two times.
- ***Custom parameters-***If there is a specific parameter that needs to be searched but there isn't a preset parameter to search, the **Custom parameter & Custom value** can be used to search for it. In this example, we are looking for records that were created by the same instrument. In the Record Detail, the information looks like this: SPE SN(spe_sn):30D . To isolate just files created by this instrument set the **Custom parameter** to **spe_sn** and the **Custom value** to **30D**.

(b) Secondary Search (red box)

While the parameters above are used for organizing the table to specific details, this search function is used to modify and search through the created table itself. To use simply type the line of text that is on the table into the box. Multiple different lines can be used as long as they are separated by a space.

For example, suppose the record in interest is known to have a peptide count of 1920, and the acquisition is known to of been made on April 4th with the extra_withdraw method recorded in the name. In this case, multiple data types can be searched for as long as they are separated by a space. This is how it would look: 1920 Apr-04 extra_withdraw



MSConnect Website: Records Page

(2) Individual Record Details

For every record recorded in MSConnect, a Record Detail page is made. Detail pages display this data in an easy-to-access way. A Record Detail can be accessed by clicking the Pk value on the table in the View Records Page. The page is organized into 3 sections to compact the data and make it easier to read. If the record is for an individual .raw file then an interactive graph will be provided.

1. Left Column

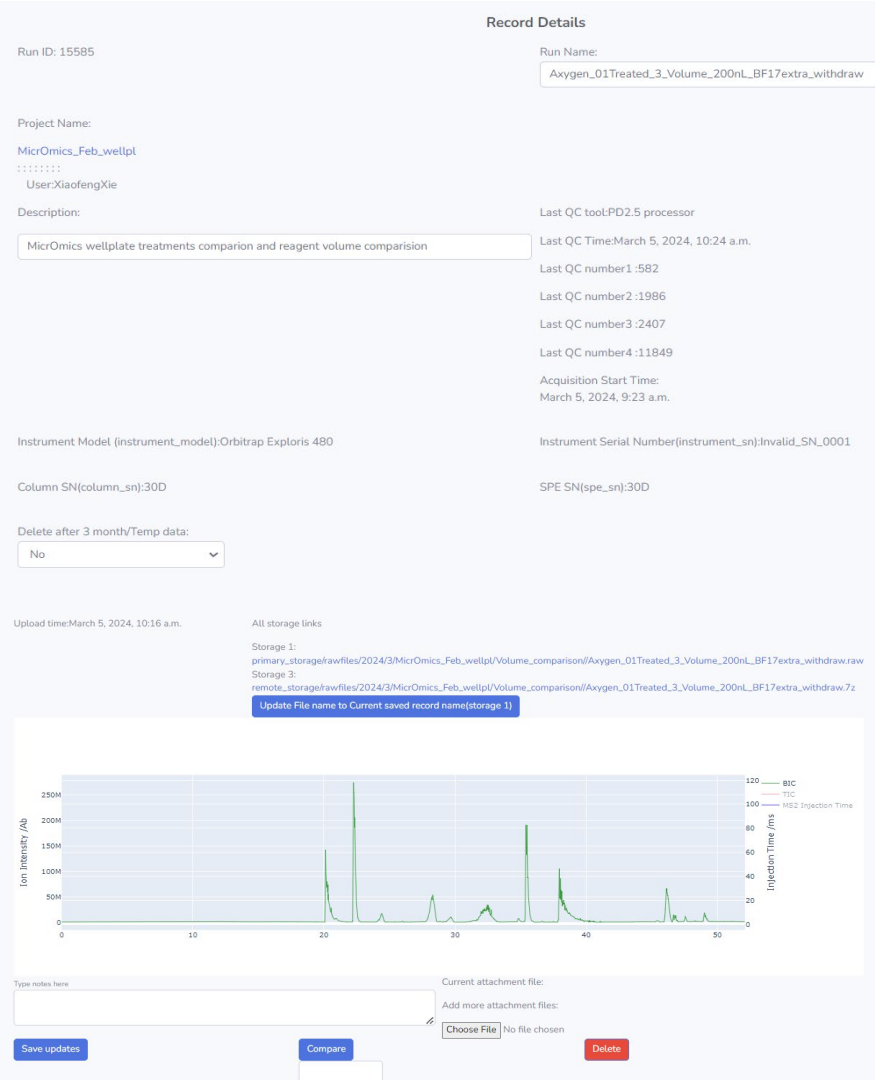
- **Run ID** - A numerical value assigned to each record in the order that they were created.
- **Project Name** - The name of the project this record is a part of.
- **User** - The user assigned to the project.
- **Description** (modifiable)- A statement describing the record was created. The Description is passed on data from the Raw file uploader.
- **Instrument Model** - The output from the machine the data was processed by.
- **Column SN** - Describes information about the analytical column.
- **Deletes after 3 Months/Temp data** (modifiable)- If set to yes, the record will be removed from the database after 3 months.
- **Upload Time** - The time when the record was uploaded

2. Right Column

- **Run Name** (modifiable)- - The name of this specific record data file.
- **QC Data** - Data created after being inserted into a processing app/wrapper. This is dependent on the type of data file that is being run.
    - **Last QC tool** - Processor used on the record.
    - **Last QC Time** - Time the processor was finished.
    - **Last QC number1** - The first data value to come from the processor, usually represents Proteins for DIA NN Processor.
    - **Last QC number2** - The second data value to come from the processor, usually represents FWHM for DIA NN Processor.
    - **Last QC number3** - The third data value to come from the processor, usually represents precursor for DIA NN Processor.
    - **Last QC number4** - The fourth data value to come from the processor, usually represents MS1 Intensity for DIA NN Processor.
- **Acquisition Start Time** - The time the Processor started.
- **Instrument Serial Number** - A serial ID that is given based on the machine the data was collected from.
- **SPE SN** - The serial number of SPE column created by the uploader
- **All Storage links** - Display where this record is stored in the storage. Clicking on the storage link downloads a copy of the file.

3. Bottom Section

- **Interactive Graph** - An interactive graph users can use to visualize the data. How to read and interact with the graph is explained in the Analyzing Interactive Graphs Diagrams Page.
- **Type Notes here** - Notes about the record can be written and saved here.
- **Save updates** - Saves all changes made to the Record Detail. To change the Run Name also in the storage, first save update, then click "Update File name" to under the storage link.
- **Add Attachments** - Allows for any file type to be uploaded and stored with the Record Detail.
- **Delete** - When clicked the record will be deleted.
- **Compare** - Insert a run ID to be compared and click compare to display the two graphs overlayed over each other.
- **Delete** - When clicked the record will be deleted.



MSconnect Website: Record Details page

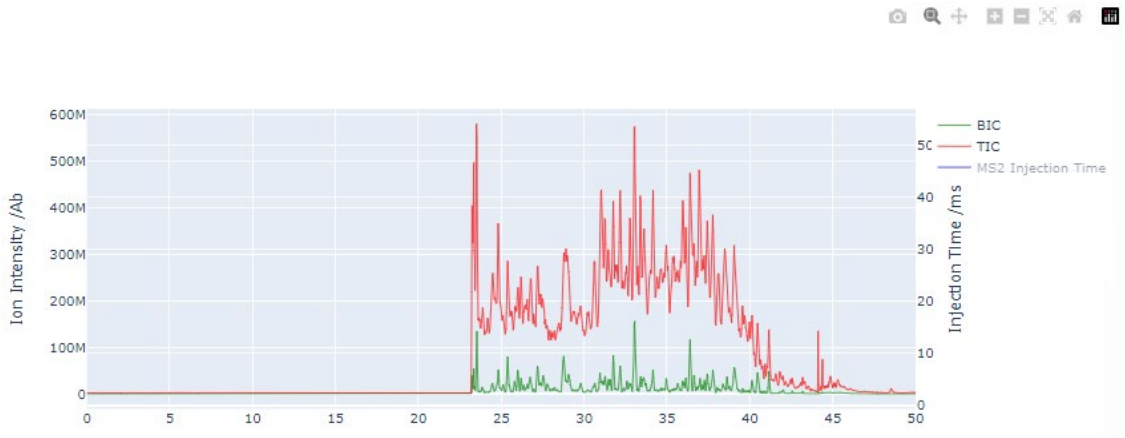
(3) Analyzing Interactive Graphs

Interactive Graphs are used by MSConnect to show .raw files in a simple way. It has three different graphs that a user can change between. .

1. Graph Types

There is an X-axis and 3 different types of Y-axes. The X-axis represents time in minutes. The Y-axis is dependent on which graph is being viewed. Click on the legend on the right to show/hide the desired graph. The graphs can also be overlayed over each other for quick and easy data comparison

- **Base ion Chromatogram (BIC)** is graphed using the Ion Intensity / Ab Y-axis
- **Total Ion Chromatogram (TIC)** is graphed using the Ion Intensity / Ab Y-axis
- **MS2 Injection Time** is graphed using the Injection Time / ms Y-axis. This is only displayed if it is a Thermo Orbitrap system.



Interactive Graph when BIC and TIC Graph Overlays

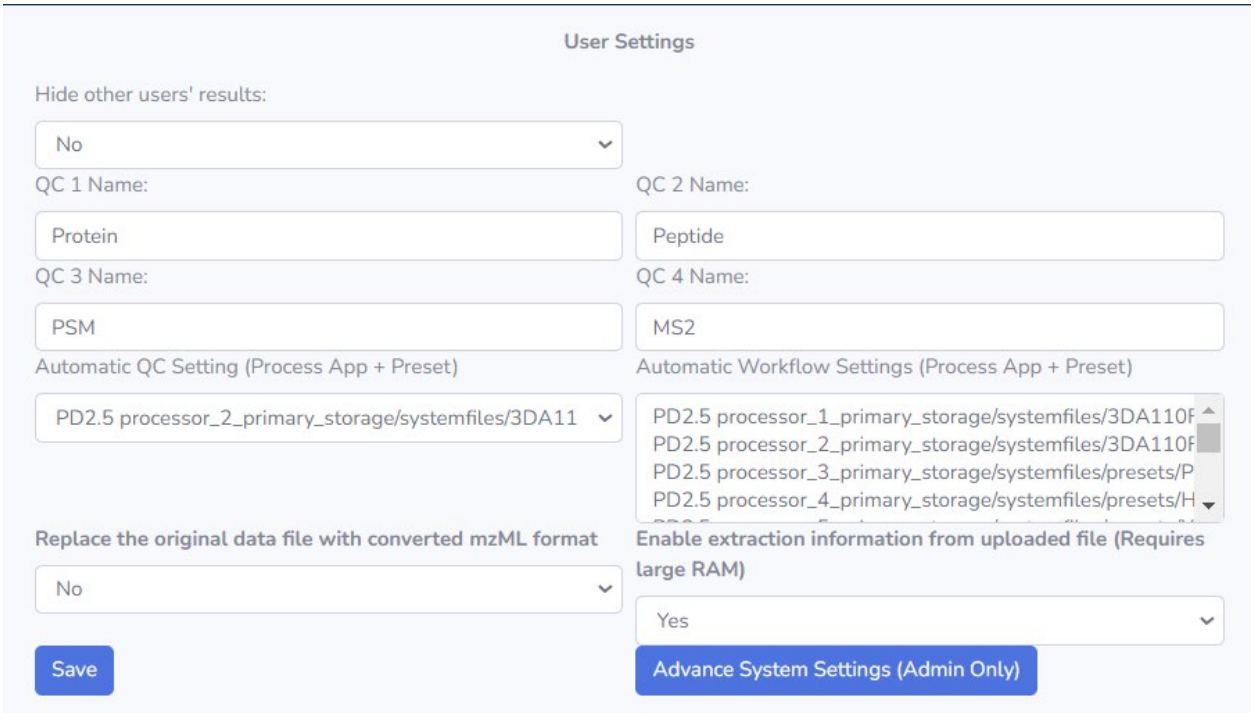
2. Tool Bar

- **Download** - Downloads the Interactive Graph to the computer
- **Zoom** - Press with the left click and drag to create a square that will become the new frame of reference.
- **Pan** - Allows the user to move the graph around by holding lift click and moving the mouse
- **Zoom in** - Zooms in entire graph based on the center point of the screen
- **Zoom out** - Zooms out the entire graph based on the center point of the screen
- **Auto scale** - Fits the graph to the screen perfectly
- **Reset axis** – Fits the graph to the screen perfectly
- **Graph provider link** - Takes user to the Interactive Graph website

e) Settings Page-

Settings are specific to each user. It also redirects to the Advanced Settings page.

The Settings page manages the settings within MSConnect specifically to a local user. Other user results can be hidden, QC headers on the Record page can be changed, and automatic processing workflow can be set. Settings can also be changed to allow interactive graphs to be created.



MSconnect Website-Setting Page

- **Hide Other users** - Displays only the current user's data and conceals all other data. (Default settings can be set in the Django Admin settings.)
- **QC Settings**
    - **QC 1 Name** - The name displayed in the record page for output_QC_number_1. The default is "Protein" for FragPipe processor
    - **QC 2 Name** - The name displayed in the record page for output_QC_number_2. The default is "Peptide" for FragPipe processor
    - **QC 3 Name** - The name displayed in the record page for output_QC_number_3. The default is "PSM" for FragPipe processor
    - **QC 4 Name** - The name displayed in the record page for output_QC_number_4. The default is "MS2" for FragPipe processor

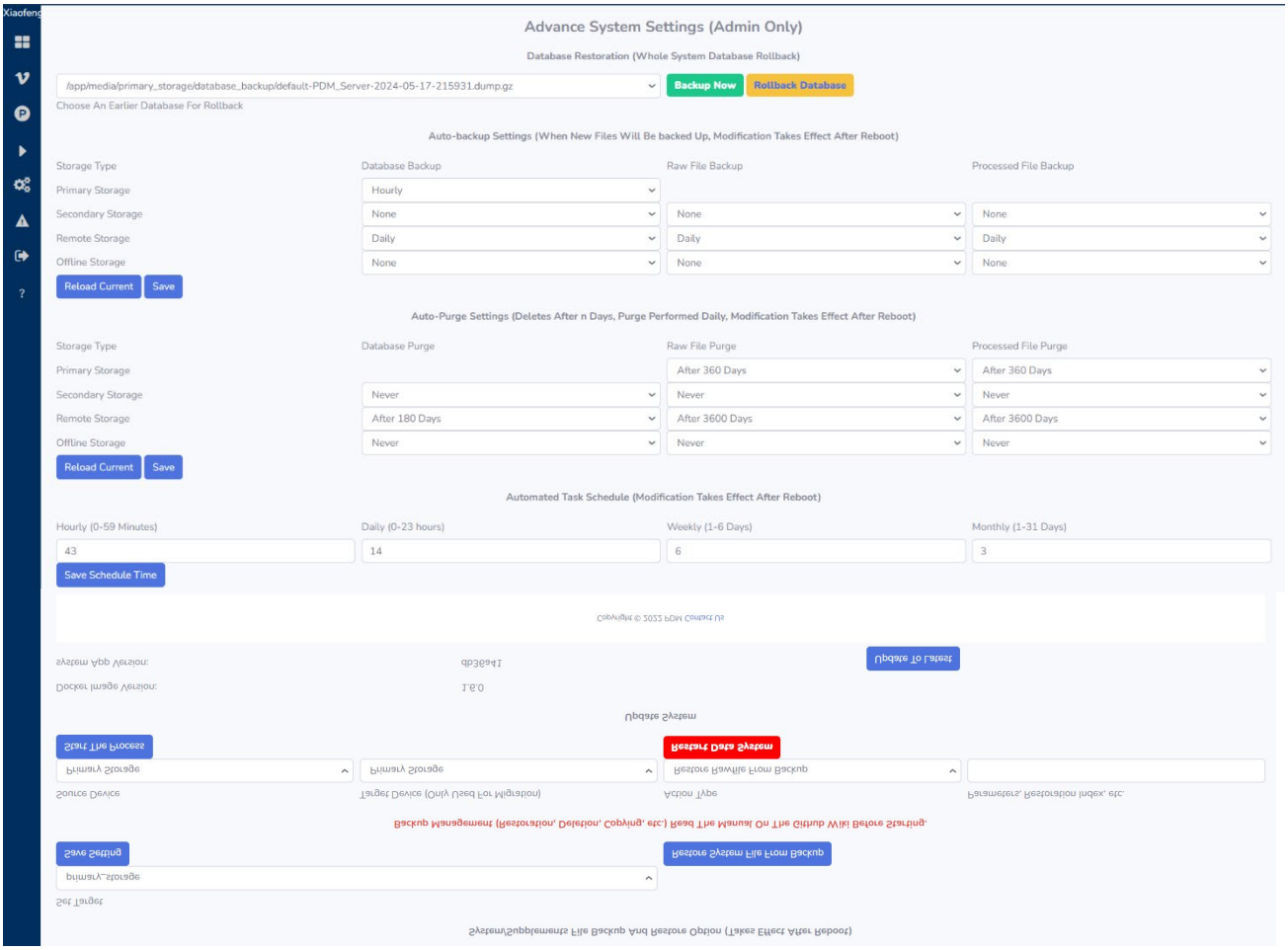
- **Automatic QC Setting** - The quality control process for individual runs is automated using the combination of the selected processing app/wrapper and preset. The selected setting will be uploaded to MSConnect along with the MS file via Raw file uploader. When set at none, the MS files will be uploaded to the storage and no automatic data processing will be carried out. When a preset is selected, the MS files will be uploaded to the storage and a processing queue will be created. Then, the processing app/worker will then pass the MS file location and the processing parameter given by preset to locally installed identification software for data procession. create a different automatic QC setting, see Create Processing workflow section.
- **Automatic Workflow Settings**- This enables a group of record files to be processed together after the last file is uploaded. To set up a group processing:
    1. Select the automatic workflow from the drop-down manual. Its default is set to none.
    2. Designate project name in Raw Files Uploader.
    3. Include “last” in the file name of the last MS files of your study.
    4. The automatic group processing will be triggered once the MS file contains “last” in the filename is uploaded to the project folder, and a queue pk number will be created.
    5. When the group processing is finished, the output file(s) can be accessed from Processing page for downstream analysis.
- **Replace Original Data with mzML format** - When turned on data is compressed into a mzML file.
- **Enable Extraction Information from Upload File** - When enabled, this allows interactive graphs to be created and displayed in every record detail file.
- **Save** - Saves all changes made to the settings
- **Advanced System Settings** - A redirect to the Advanced System Settings page where data management settings are modified.

(1) Advanced system settings (Admin)

The Advanced System Settings page is important for MSConnect program management. Settings involving data management and backups will be found on this page.

- **Database Restoration -** Based on the set auto backup settings backups are created regularly. They can be accessed here. A backup can be selected and by clicking "Rollback Database" the entire database is returned to this backup. A backup is created by clicking "Backup Now"
- **Auto-backup Settings -** The backup setting backups the Database, Raw Files, and Processed Files. There are also settings for each of the four storage types; Primary Storage, Secondary Storage, Remote Storage, and Offline Storage.

- **Auto-Purge Settings -** To prevent an overabundance of data the auto-purge setting deletes data after the set amount of time. It follows the same organization as the data backup setting.
- **Automated Task Schedule** - In the Auto-Backup there are 5 options; Hourly, Daily, Weekly, Monthly, and None. The Automated task schedule matches these values. If Hourly is selected and hour is set to 43, every 43 minutes a backup will be created.
- **System/Supplements File Backup And Restore Option**
- **Backup Management**
- **Update Program -** Clicking "Update To Latest" searches for an update for the program and installs it.



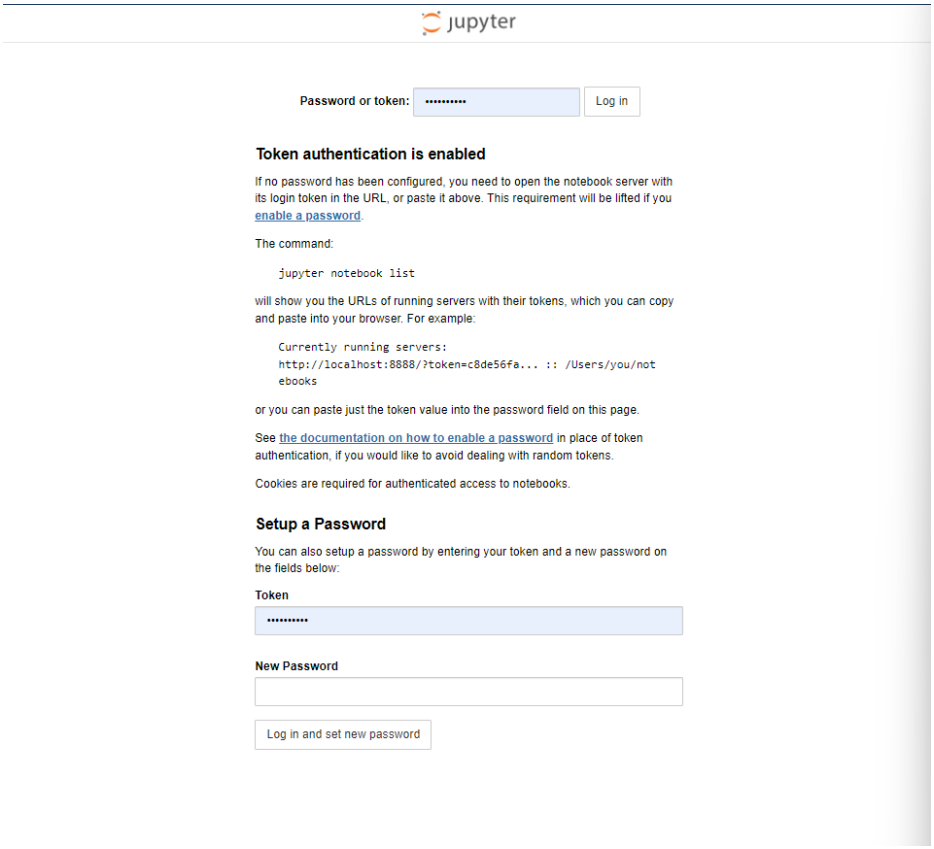
MSconnect Website: Advance System Setting Page

f) Advanced page (developer)- connects MSConnect to jupyter.

This page connects MSConnect to jupyter. Clicking this tab redirects the user to Jupyter. Jupyter notebook let you write you own Python script and access all the data the MSConnect has. It is useful for developer or very advanced user. However, it can be risky, so it is protected by a single password.

The password is set in the installation file. To find the Jupyter password. First, access the MSConnect program source files. Open the file called ".django_secrets.env", there is a setting called JUPTER_PASSWORD that contains the password. it looks like this:

JUPTER_PASSWORD ='itsasecret'



MSConnect Website: Advanced Page

g) Logout icon- Logs the user out upon clicking

h) Help Page- Hub for other links

The Help page contains various links that lead the user to helpful content

- **Signup for new account** - Reroutes the user to the Registration page
- **Admin** - Links the user to the Admin page
- **Django Logs**- Reroutes the user to the Logs page, if the user is an admin
- **Download Lastest Upoader** -To download the latest Raw Files Uploader. The raw file uploader is used to upload data to MSConnect
- **Wiki**-Links to the User manual on Wiki of MSConnect Github page
- **User Community**- Reroutes the user to a community page where MSConnect-related discussions are held (under construction)

(1) Admin Page

1. How to Access

To access the Django admin page, click on the help tab, then the admin link. This can only be accessed if logged in as an admin.

2. Authentication and Authorization

These settings provide permissions for the users. The group setting is used to group permissions. A user can be added to a group to receive those permissions.

- **Groups** - Groups exist to provide easy command access to users. On this page, groups can be created/deleted and given permission. When a user is added to a group, they receive all the permissions of that group.
- **Users** - This page shows user data. Every created user is displayed here. Users can be added or deleted. Clicking on a user allows the admin to modify information about the user, this includes username and other important user information.

3. File Manager

- **Data Analysis queues** - It is used to manage data processing queued. The queue pk number is used here.
- **File Storages** - For each MS file uploaded to MSConnect, a FileStorage object is created to store the file location.
- **Processing apps** – This use to manage established and create new automatic QC Setting preset under Setting Page

- **Sample Records** - This is where individual uploaded record files are stored
- **Saved Visualizations** - Shows information on visualization tools
- **System Settings** - Holds system settings from the Advanced System Settings page.
- **User Settings** – Under Construction
- **Visualization Apps** - There is a "visualization apps" setting for each visualization tool. A visualization app object holds the settings for each visualization tool.
- **Worker Status** - This page stores information on each processing worker that is installed.

VI. FAQ

1. How do I upload a MS file to MSConnect?

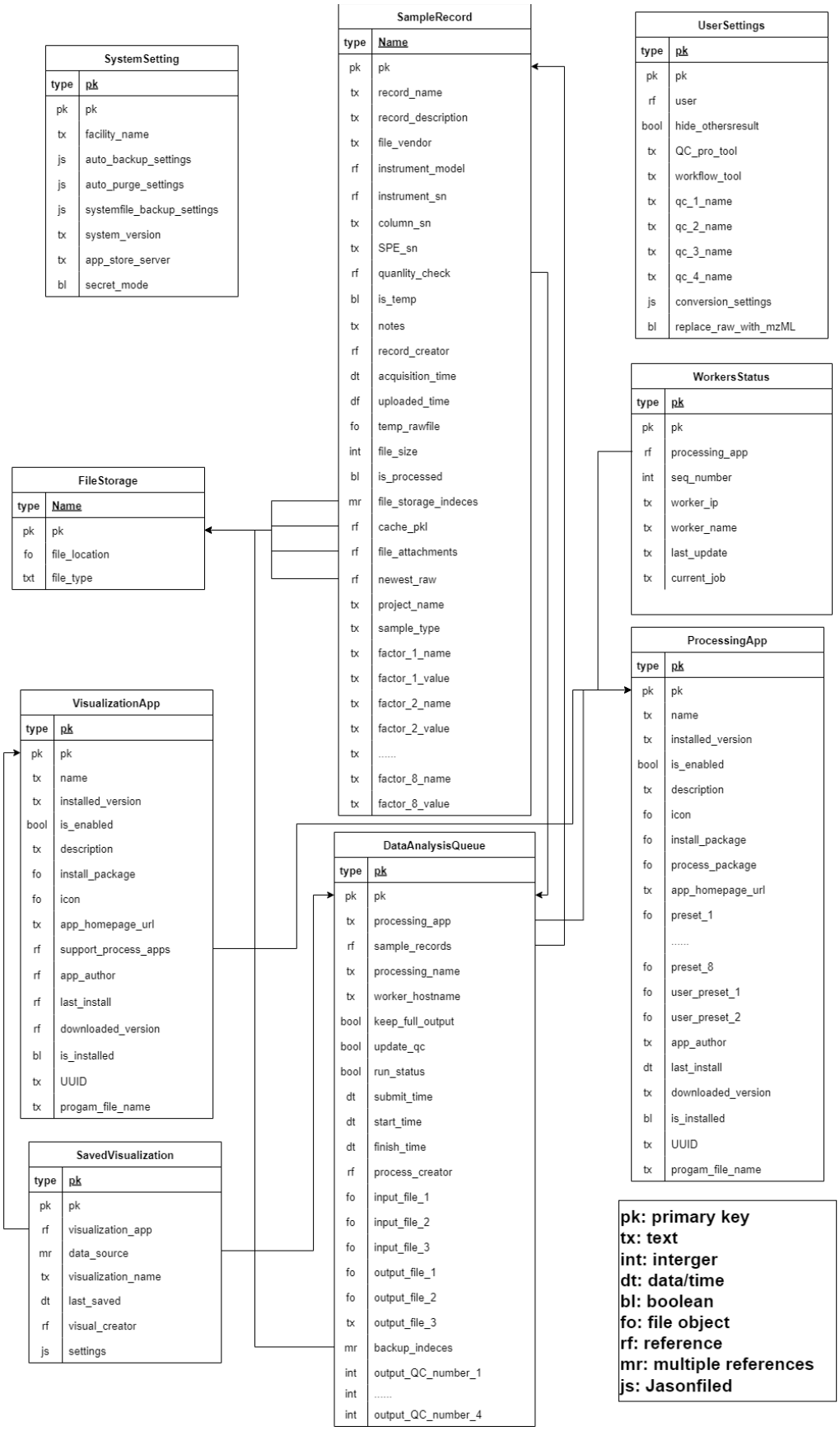
Follow the instructions on the Raw File Uploader Github or follow the instructions here

2. How can I find more support for using MSConnect?

You are welcome to contact us by emailing to **MSConnect@gmail.com**, or submit a ticket on the MSConnect GitHub site. You can also find the YouTube tutorial in our YouTube Channel: https://www.youtube.com/channel/UCahxAV0zD4LCYiL71-UNeAA

3. What is Database Schema and Data Structure?

The data structure of Django is defined by models.py file inside web/file_manager/ folder. Details on how to create queries can be found in the official Django guide. These queries can be executed in the main server app, processing applications, visualization apps, and Jupyter notebooks.


SystemSetting
type pk
pk pk
tx facility_name
js auto_backup_settings
js auto_purge_settings
js systemfile_backup_settings
tx system_version
tx app_store_server
bl secret_mode
SampleRecord
type Name
pk pk
tx record_name
tx record_description
tx file_vendor
rf instrument_model
rf instrument_sn
tx column_sn
tx SPE_sn
rf quanlity_check
bl is_temp
tx notes
rf record_creator
dt acquisition_time
df uploaded_time
fo temp_rawfile
int file_size
bl is_processed
mr file_storage_indeces
rf cache_pkl
rf file_attachments
rf newest_raw
tx project_name
tx sample_type
tx factor_1_name
tx factor_1_value
tx factor_2_name
tx factor_2_value
tx .....
tx factor_8_name
tx factor_8_value
UserSettings
type pk
pk pk
rf user
bool hide_othersresult
tx QC_pro_tool
tx workflow_tool
tx qc_1_name
tx qc_2_name
tx qc_3_name
tx qc_4_name
js conversion_settings
bl replace_raw_with_mzML
WorkersStatus
type pk
pk pk
rf processing_app
int seq_number
tx worker_ip
tx worker_name
tx last_update
tx current_job
FileStorage
type Name
pk pk
fo file_location
txt file_type
ProcessingApp
type pk
pk pk
tx name
tx installed_version
bool is_enabled
tx description
fo icon
fo install_package
fo process_package
tx app_homepage_url
fo preset_1
.....
fo preset_8
fo user_preset_1
fo user_preset_2
tx app_author
dt last_install
tx downloaded_version
bl is_installed
tx UUID
tx progam_file_name
VisualizationApp
type pk
pk pk
tx name
tx installed_version
bool is_enabled
tx description
fo install_package
fo icon
tx app_homepage_url
rf support_process_apps
rf app_author
rf last_install
rf downloaded_version
bl is_installed
tx UUID
tx progam_file_name
DataAnalysisQueue
type pk
pk pk
tx processing_app
rf sample_records
tx processing_name
tx worker_hostname
bool keep_full_output
bool update_qc
bool run_status
dt submit_time
dt start_time
dt finish_time
rf process_creator
fo input_file_1
fo input_file_2
fo input_file_3
fo output_file_1
fo output_file_2
tx output_file_3
mr backup_indeces
int output_QC_number_1
int .....
int output_QC_number_4
SavedVisualization
type pk
pk pk
rf visualization_app
mr data_source
tx visualization_name
dt last_saved
rf visual_creator
js settings
pk: primary key
tx: text
int: interger
dt: data/time
bl: boolean
fo: file object
rf: reference
mr: multiple references
js: Jasonfiled

VII. Troubleshooting Guide

1. Installation Error: *.django_secrets.env: The system cannot find the file specified.*

```
Failed to load C:\Users\kelec\MS Connect\Proteomic-Data-Manager\.django_secrets.env: open C:\Users\kelec\MS Connect\Proteomic-Data-Manager\.django_secrets
.env: The system cannot find the file specified.

 *  The terminal process "C:\WINDOWS\System32\WindowsPowerShell\v1.0\powershell.exe -Command docker compose -f "docker-compose.yml" up -d --build" termina
ted with exit code: 1.
 *  Terminal will be reused by tasks, press any key to close it.
```

If you are getting this error in yellow, your .django_secrets.env file is not recognized by docker. Make sure to name the file properly.

2. Installation Error: *.django_secrets.env: The system cannot find the path specified.*

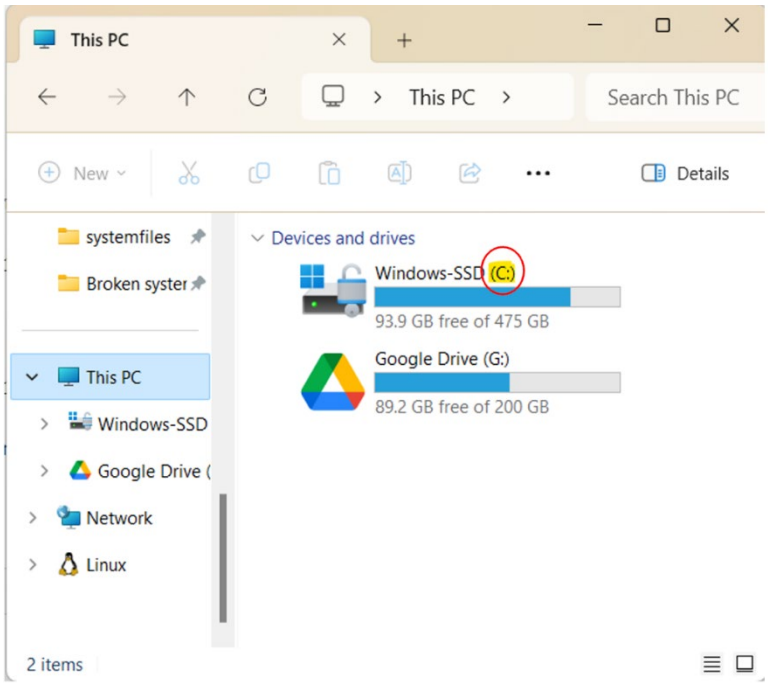
```
[+] Running 0/0
 - Container proteomic-data-manager-data_manager-JupiterNotebook-1  Recreate                                                                    0.0s
 - Container proteomic-data-manager-data_manager-django-1           Recreate                                                                    0.0s
Error response from daemon: mkdir D:: The system cannot find the path specified.

 *  The terminal process "C:\WINDOWS\System32\WindowsPowerShell\v1.0\powershell.exe -Command docker compose -f "docker-compose.yml" up -d --build" termina
ted with exit code: 1.
 *  Terminal will be reused by tasks, press any key to close it.
```

If you are getting this error in yellow, the file path in your docker-compose.yml file could be incorrect. To fix this,

a. Open the docker-compose.yml and scroll down to ~line 50. you should see something like this:

```
50        - type: bind
51          source: /D/docker/primary_storage # primary storage, default saving everything in C drive root, please change
52          target: //app/media//primary_storage #don't change
53        #secondary_storage, optional but recommended
54        - type: bind
55          source: /D/docker/secondary_storage
56          target: //app/media//secondary_storage #don't change
57        #remote_storage, optional but recommended
58        - type: bind
59          source: /D/docker/remote_storage
60          target: //app/media//remote_storage #don't change
61        #offline_storage, optional but recommended
62        - type: bind
63          source: /D/docker/offline_storage
64          target: //app/media//offline_storage #don't change
```

b. Now, make sure your drivers match the file path. To check what kind of drive you have, open any type of folder and click on "This PC"



c. Find out which drive you want to store your data on and change the /D/ to that driver. For example, mine would look like this

a

```
      - type: bind
        source: /C/docker/primary_storage # primary storage, default saving everything in C drive root, please change
        target: //app/media//primary_storage #don't change
      #secondary_storage, optional but recommended
      - type: bind
        source: /C/docker/secondary_storage
        target: //app/media//secondary_storage #don't change
      #remote_storage, optional but recommended
      - type: bind
        source: /C/docker/remote_storage
        target: //app/media//remote_storage #don't change
      #offline_storage, optional but recommended
      - type: bind
        source: /C/docker/offline_storage
        target: //app/media//offline_storage #don't change
```