# Data Cleanup Toolset

# Quick Start Guide

Oregon Department of Transportation

Prepared by

Engineering Applications Support Team

**Data Cleanup Toolset–simple tools to help clean up your files.** The Data Cleanup Toolset will help ODOT employees manage their F:\ drives and shared server locations such as a Crew Share. Regular use of the toolset will help ensure good data management practices are followed.

The Data Cleanup Toolset consists of three utilities that aid in cleaning up storage space with these functions:

- Find files/folders that take up the most space
- Find duplicate files
- Find paths longer than 200 characters

The toolset can be installed as a bundle by any ODOT user. See the installation instructions at the end of this document **(jump to the end for installation instructions.)** You can also contact the Computer Support Desk or the DMV IS Technical Resource Center for installation assistance.

**WinDirStat – find large files –** This is a visually friendly disk usage statistics viewer to find large files. Use this program first to get a grasp of how much storage you are using and where to prioritize the cleanup of your disk space **(jump to page 4 for instructions for finding large files.)**

**SearchMyFiles – find duplicate files –** This general purpose search tool can be used to find duplicate files. Use this program to find duplicate files on your F:\ drive and crew share. Duplicate files should be deleted from a server location **(jump to page 10 for instructions for finding duplicate files.)**

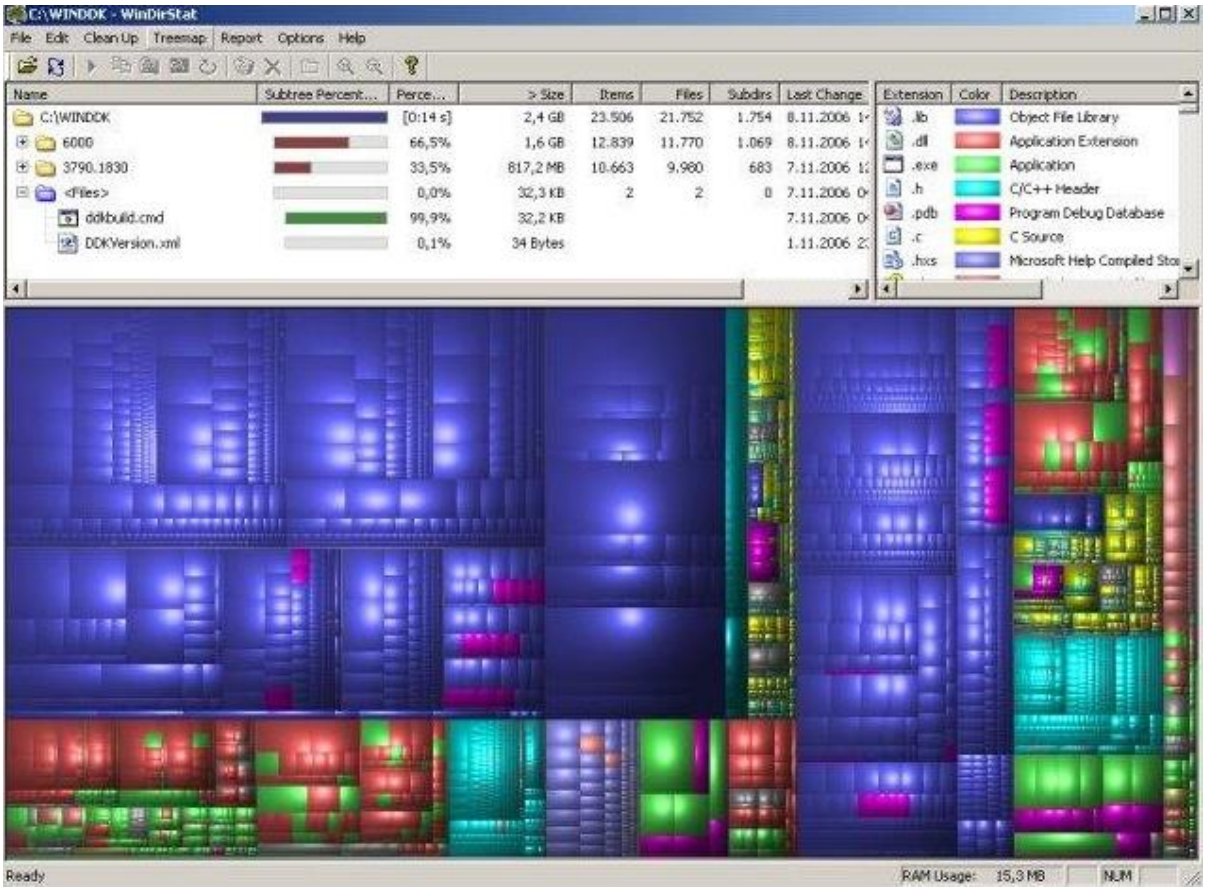
**Path Scanner – find long paths –** Use this tool to scan your network storage such as your F: drive and crew share to find paths which are longer than 200 characters. Long file paths don't use a significant amount of storage space. However when the path length is too long, the files are not reliably backed up which increases the chance that you could lose your data! This hazard can easily be avoided by checking your data with Path Scanner **(jump to page 16 for instructions on finding long paths.)**

**Long Term File Storage and Data Management Resources –** Here are some resources that may help you manage your data.

| ***Division*** | **Description** | **Details** |
|---|---|---|
| Agency Wide | ODOT Records Clean Out Day | The ODOT Records Clean Out Day: The Five Essential Elements of Records Cleanout Day document includes records retention, litigation holds, and other cleaning tips.<br>Business Services has other record management resources. |
| Agency Wide | Data Management | The Computer Support Desk has some data management resources. See the **Other Resources** menu on the right side of the Computer Support Desk home page.<br>Direct Data Management links<br>• Best Practices<br>• Concepts<br>• Issue: Long File/Folder Names |
| Agency Wide | Crew Share | Data read or used by multiple members of a crew may be stored at a single location on a crew share drive rather than each employee stores their own copy. |
| Highway Division | Engineering Archive | Engineering Archive Requests may be made for engineering project data. |
| Highway Division | ProjectWise | As ProjectWise comes on-line, construction engineering, and project data will be stored here. |
| Highway Division | Highway Archive | The Highway Division has data management resources for non-engineering data. |
| | | |
| | | |

# WinDirStat

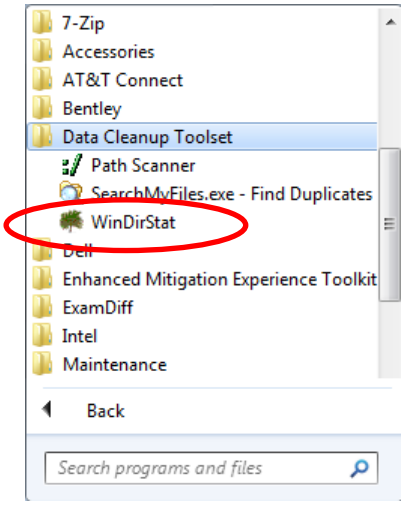
# Quick Start Guide




## WinDirStat - Find large files

WinDirStat (Windows Directory Statistics) is a disk usage statistics viewer and cleanup tool for Windows. It is used to find large files and folders with a large volume of files illustrated with colored rectangles that are proportional to the file's size. The colored rectangles aid in quick identification of files taking up the most drive space.

## Launching WinDirStat




When the application is installed, WinDirStat may be launched from the icon found in the Data Cleanup Toolset folder from the Start menu.

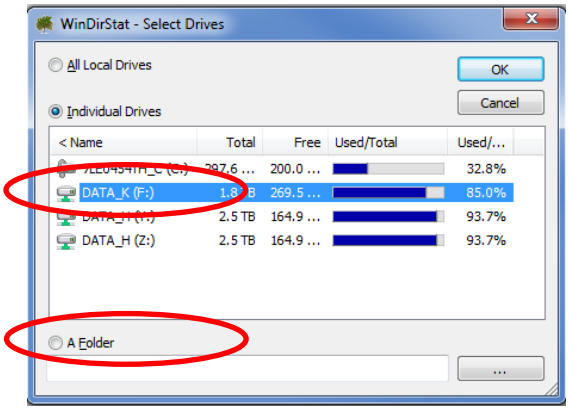
Using the mouse, click on the Windows icon (Start button) , click on **All Programs**, scroll to and click on the **Data Cleanup Toolset** folder.

You will see all three available tools. Click on the **WinDirStat** icon to launch the application.

## Using the WinDirStat

### Selecting the Drive to Scan

When the application is initially launched, you will be prompted to select a drive (if you have multiple hard drives or mapped network drives) as shown below. Typically you will select the **Individual Drives** radio button and the **F: drive**. Then click on the **OK** button or use the Enter key on the keyboard. You may see an animated icon (Pac-Man like) moving back and forth as the drive is being scanned.

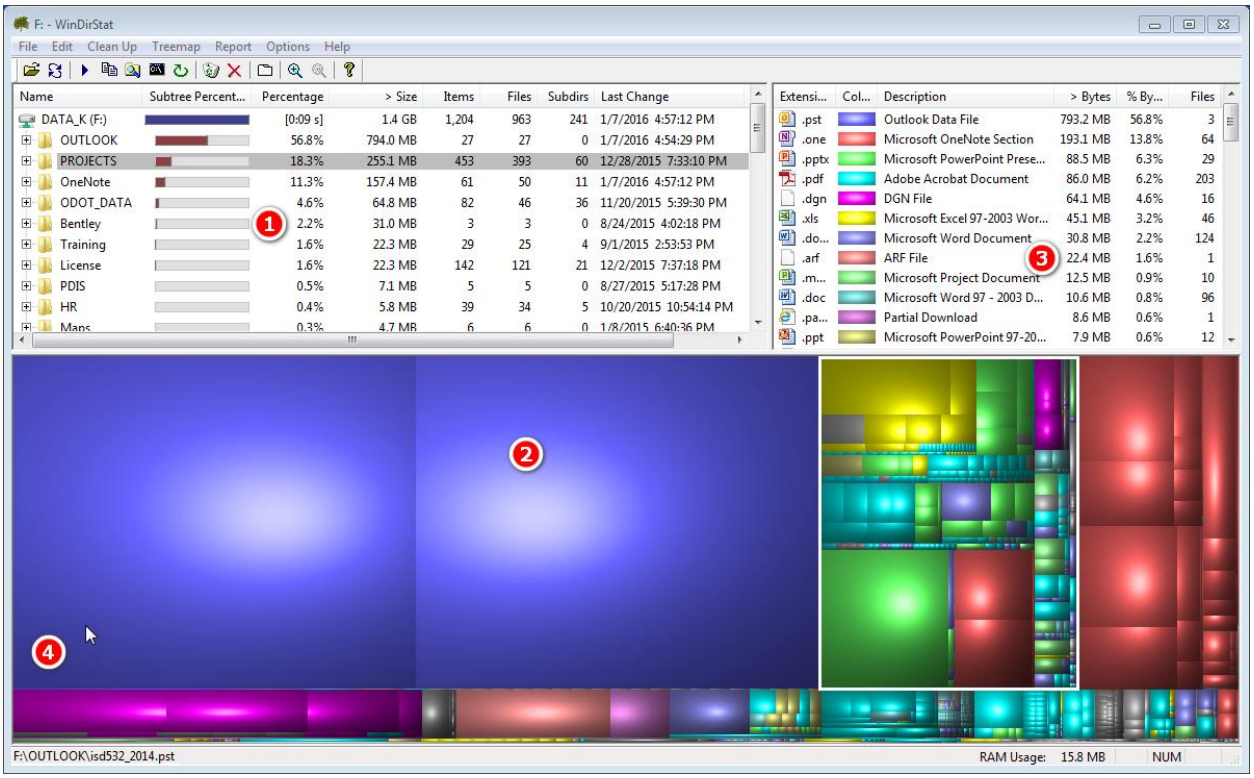



Scanning may be limited to a specific folder on a drive or crew share by choosing the **A Folder** radio button at the bottom of the Select Drives dialog box and navigating to the folder of interest.

## Navigating the Treemap

WinDirStat reads the all of the folders once and then presents it in three useful views:

- The Directory List (1) resembles the folder view of the Windows Explorer by default is sorted by file/subtree size, note the "> Size" column header
- The Treemap (2) shows the whole contents of the folders represented by colored rectangle
- The Extension List (3) serves as a legend and shows statistics about the file types
- The file name is displayed in the lower left corner (4) when a file/folder is selected or when the mouse hovers over one of rectangles

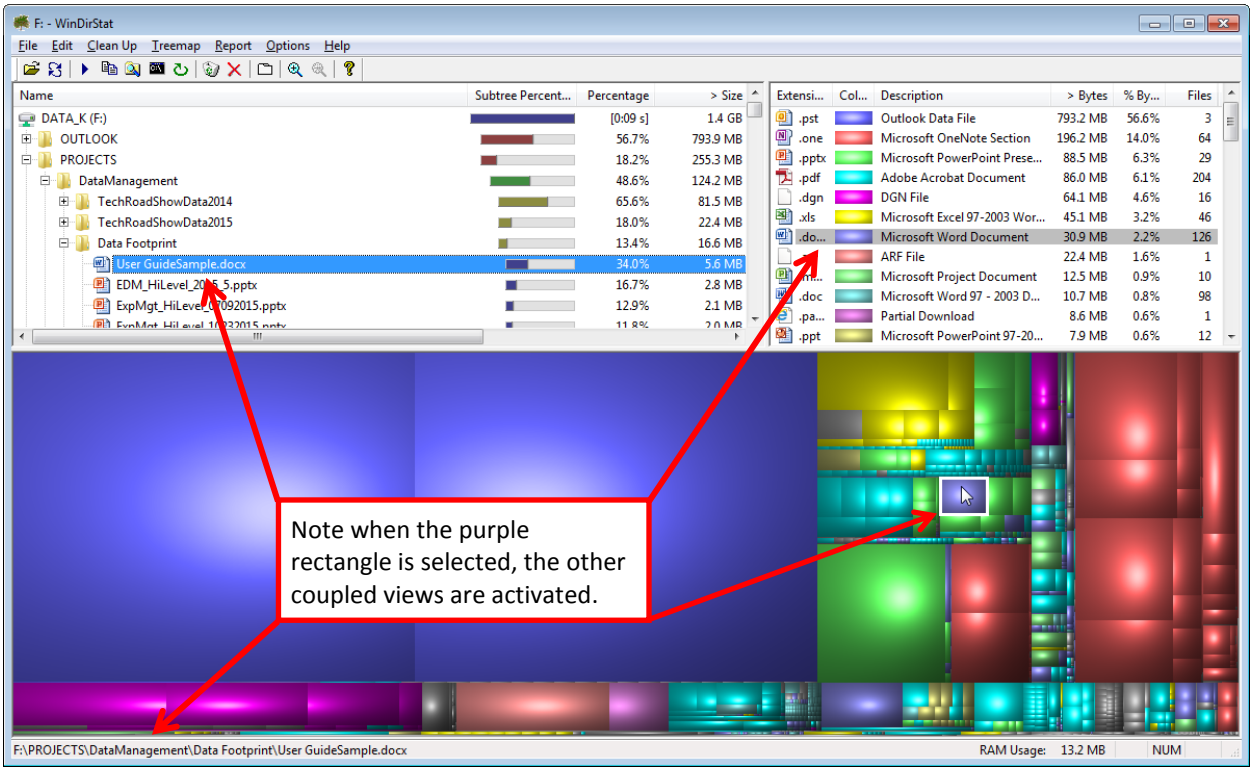



Like Windows Explorer, the Directory List folders can be expanded by double clicking on the folder name. Unlike Windows Explorer, the folders and file are also listed together with the largest files at the top. The upper part of the list is where you should focus your attention for managing the space.

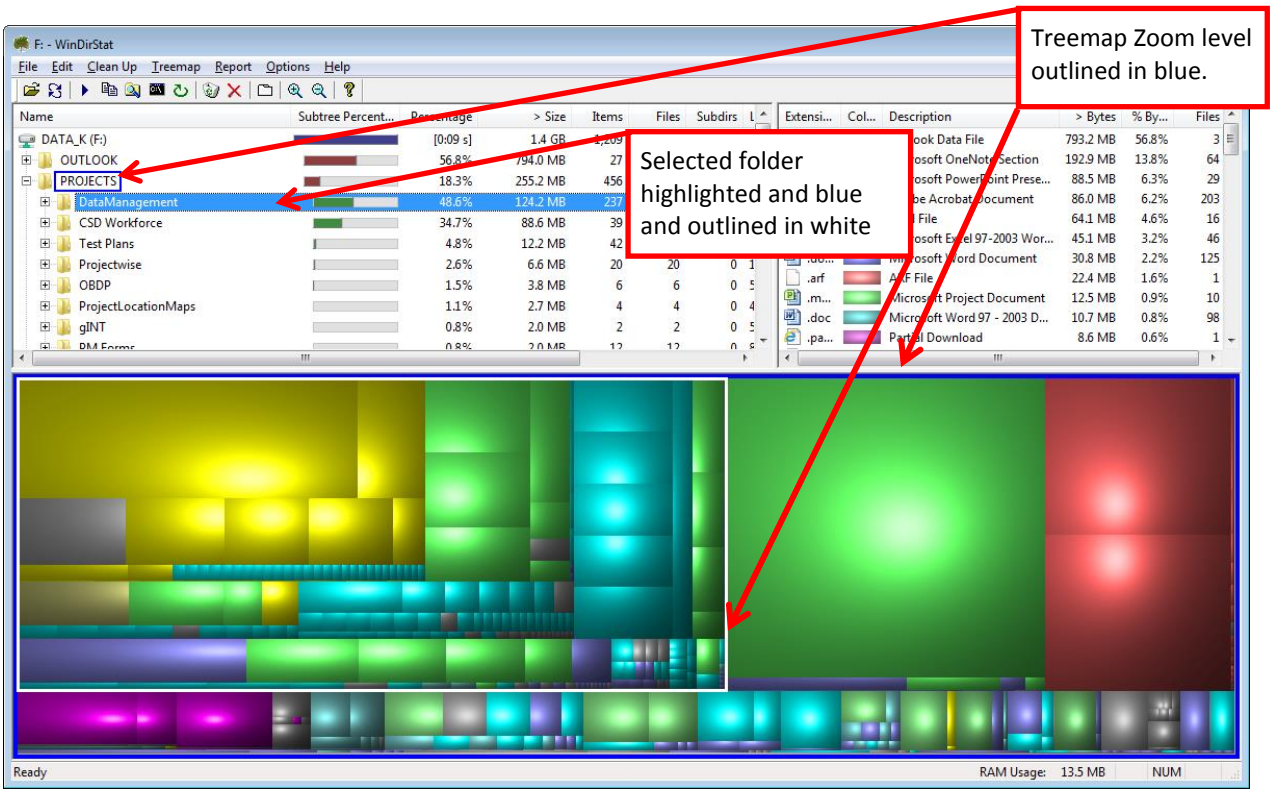
The Treemap colored rectangles are proportional to the file's size. The rectangles are arranged in such a way, that folders make up rectangles, which contain all their files and sub-folders. The area of the rectangles is proportional to the size of the subtrees. The color of a rectangle indicates the type of the file, as shown in the extension list. The large rectangles and large groups of rectangles is where you should focus your attention for managing the space.

## Using the Directory List and Treemap Together

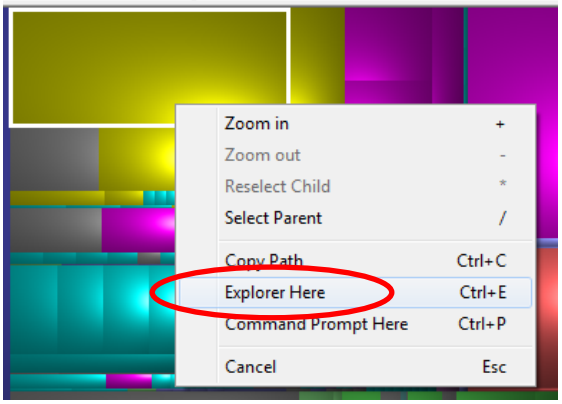
The Directory List and Treemap views are coupled together. When a folder or file is selected, it is also selected in the other view. In the previous screenshot, the PROJECTS folder is selected in the Directory List and is also outlined in a white box in the Treemap. In the following screenshot the purple rectangle for file UserGuideSample.docx is selected in the Treemap view as indicated by the white outline and is also selected in the Directory List with the blue highlight.



You may want to zoom in on a Treemap view of a folder if there are a lot files. To zoom in, first select a subfolder in the Directory List. In the example below the Data Management folder is selected. Next select **Treemap > Zoom In.** The **+** on the numeric keypad or the icon also work. The displayed Treemap folder shifts one folder level. The displayed level is outlined in blue in both the Directory List and Treemap view. In the example below the Treemap changed from F:\ to PROJECTS. Selecting the Zoom In command a second time would shift to the Data Management folder. The Zoom Out command will shift one folder level back out.



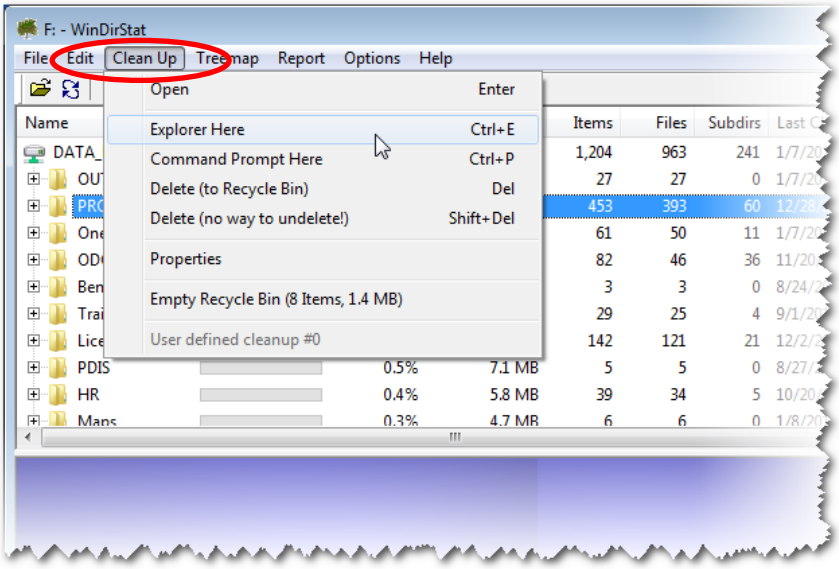
## Managing the Space




Once the file(s) have been identified, you can choose how to manage the file appropriately to optimize storage space availability. For example, left click once to highlight a specific rectangle. In the picture below, the yellow rectangle represents an Excel (.xlsx) file. A right click of the mouse opens a menu with several options. The **Explorer Here** option will open Windows Explorer to manage files outside of the WinDirStat environment.

## Clean Up

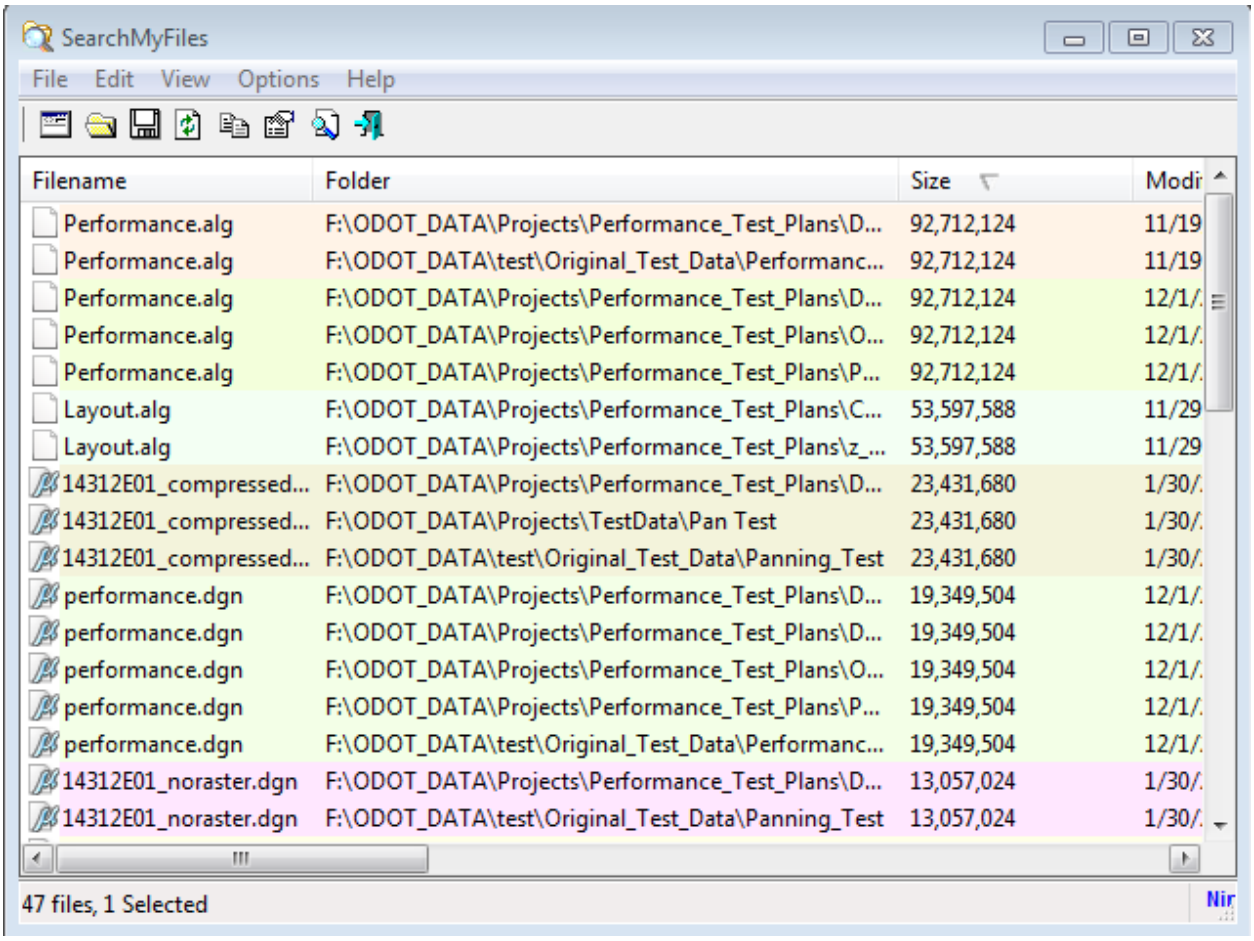
Another option for file management is highlighting a rectangle and selecting an option from the main toolbar at the top of the application window. In this example, the picture illustrates the **Clean Up** menu options.




***Warning:*** *Do NOT delete any file(s) unless you are certain the file is a duplicate and the original is stored in another archived or protected location, and not subject to any required records retention rules. To avoid the risk of data loss, always check with the appropriate data owner(s) prior to deleting files you did not create, files from Crew Share drives, or any shared file locations.*

# SearchMyFiles

# Quick Start Guide



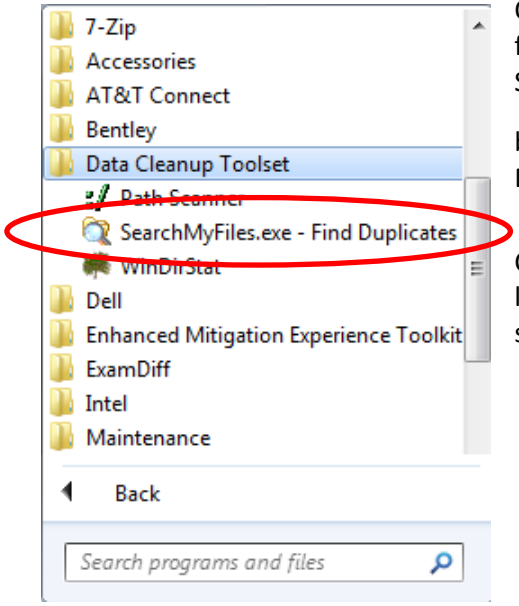
## SearchMyFiles - Find duplicate files

SearchMyFiles is an alternative to the standard Windows file search tool. It can quickly identify duplicate files taking up the most drive space.

For data cleanup, a SearchMyFiles - Find duplicate files shortcut has been set up with the parameters set to find duplicate files on the F: drive.

Windows Explorer is recommended for basic file searches. For advanced general file searches, SearchMyFiles is available but it is not covered in this documentation.

## Launching SearchMyFiles




Once the application is installed, SearchMyFiles may be launched from the icon found in the Data Cleanup Toolset folder from the Start menu. Using the mouse, click on the Windows icon (Start button) , click on **All Programs**, scroll to and click on the Data Cleanup Toolset folder. You will see all three available tools.

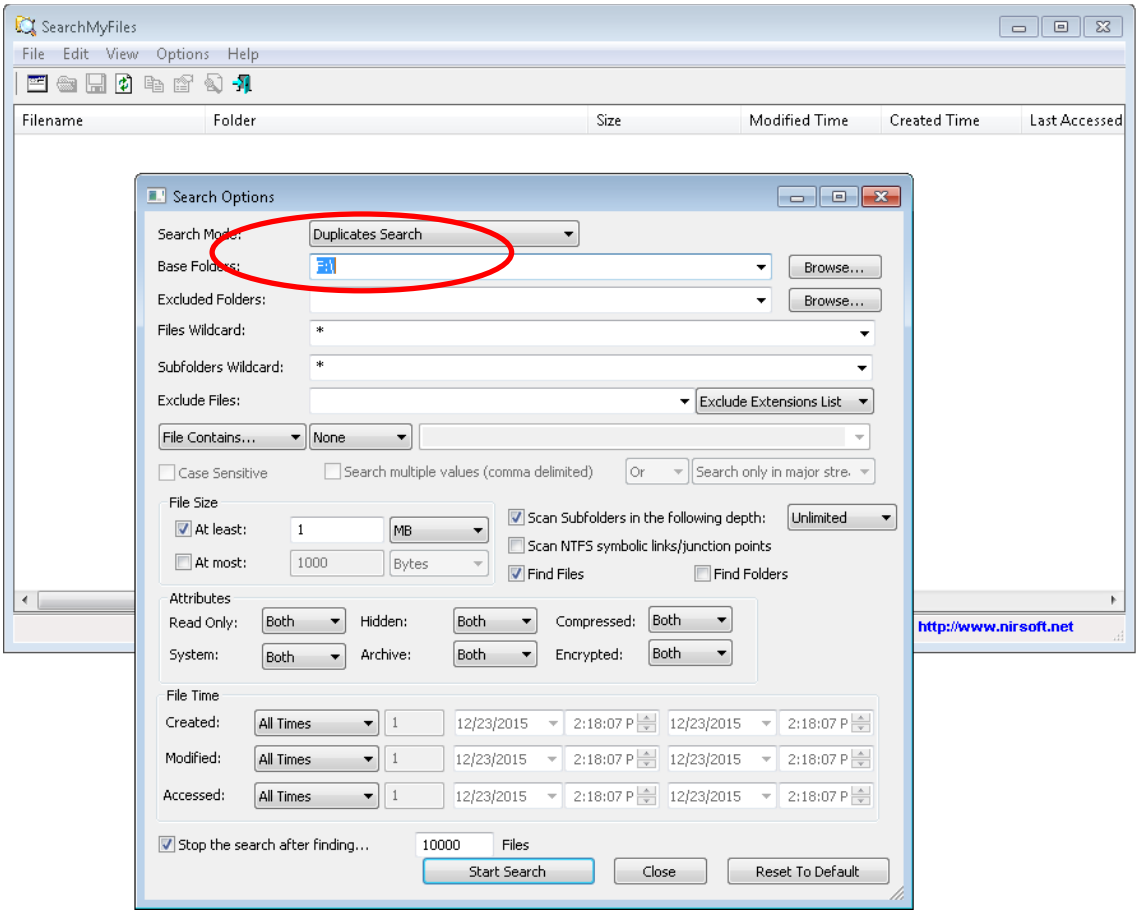
Click on the **SearchMyFiles.exe – Find Duplicates** icon to launch the application. This will open the initial two windows as seen in the following screenshots.

## Using the Application

When the application is launched with the **SearchMyFiles.exe – Find Duplicates,** the application will default to show 'Duplicates Search' in the **Search Mode** dropdown field and '**F:**' in the **Base Folders** field. You may **Browse…** to a different drive or folder path or manually type in a different valid path.

A second location like a crew share can be included as a Base Folder separated by a semi-colon. For example, the Base Folders: section could show F:\; CrewShare path. The search would include both locations. Next, click on the **Start Search** button to begin the search for duplicate files.

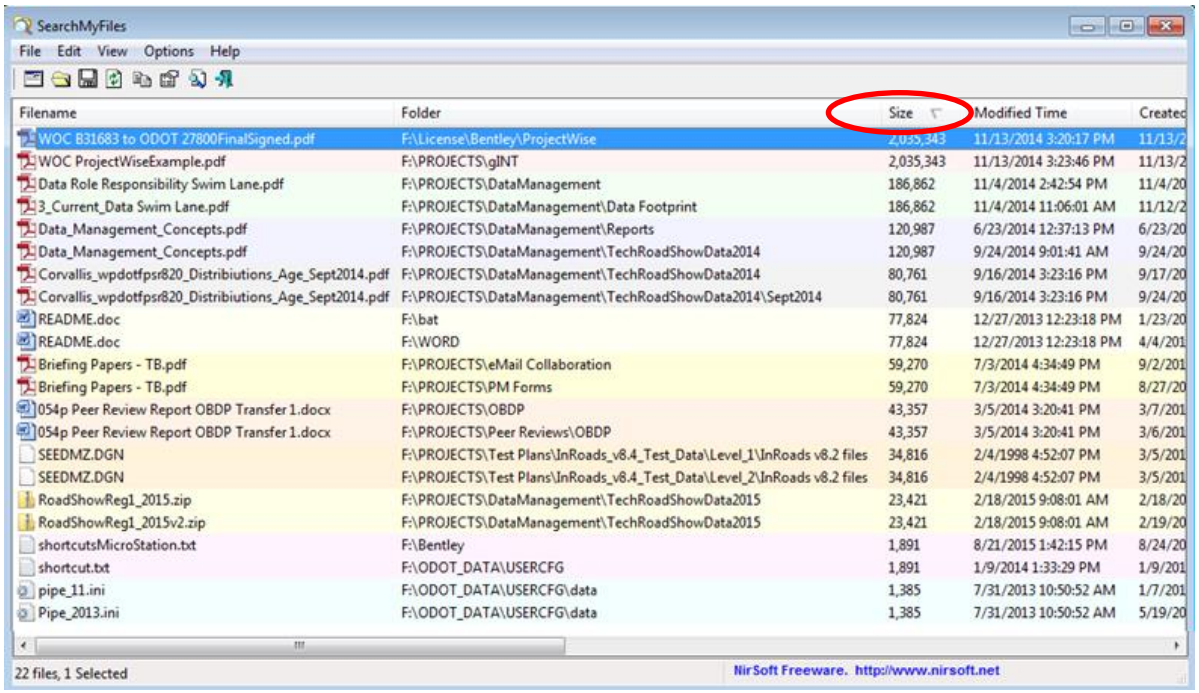
By default, only files 1 MB or larger will be reported.




The Duplicates Search mode is the primary intended use for this application. However, there are options to use the Wildcard (*) character and a partial name to perform targeted searches. Also, the **File Time** section options allow for very specific searches based on date ranges.

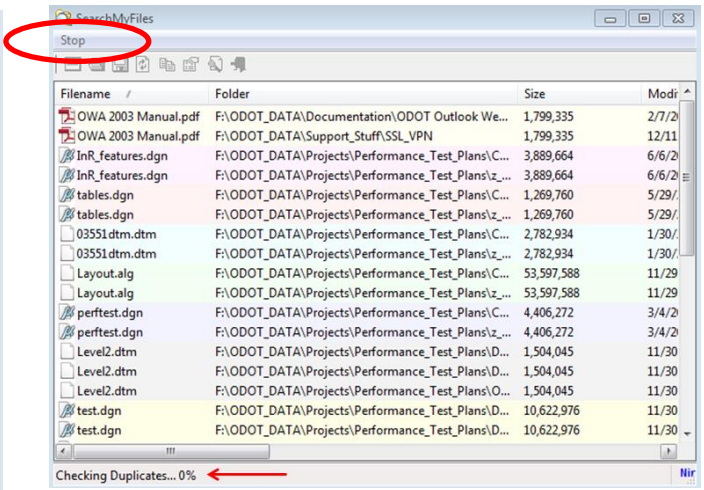
## Reading the Results

Once the search has completed the results will be displayed in the **SearchMyFiles** main window. Click on the **Size** column header to sort the list by largest to smallest duplicate files.

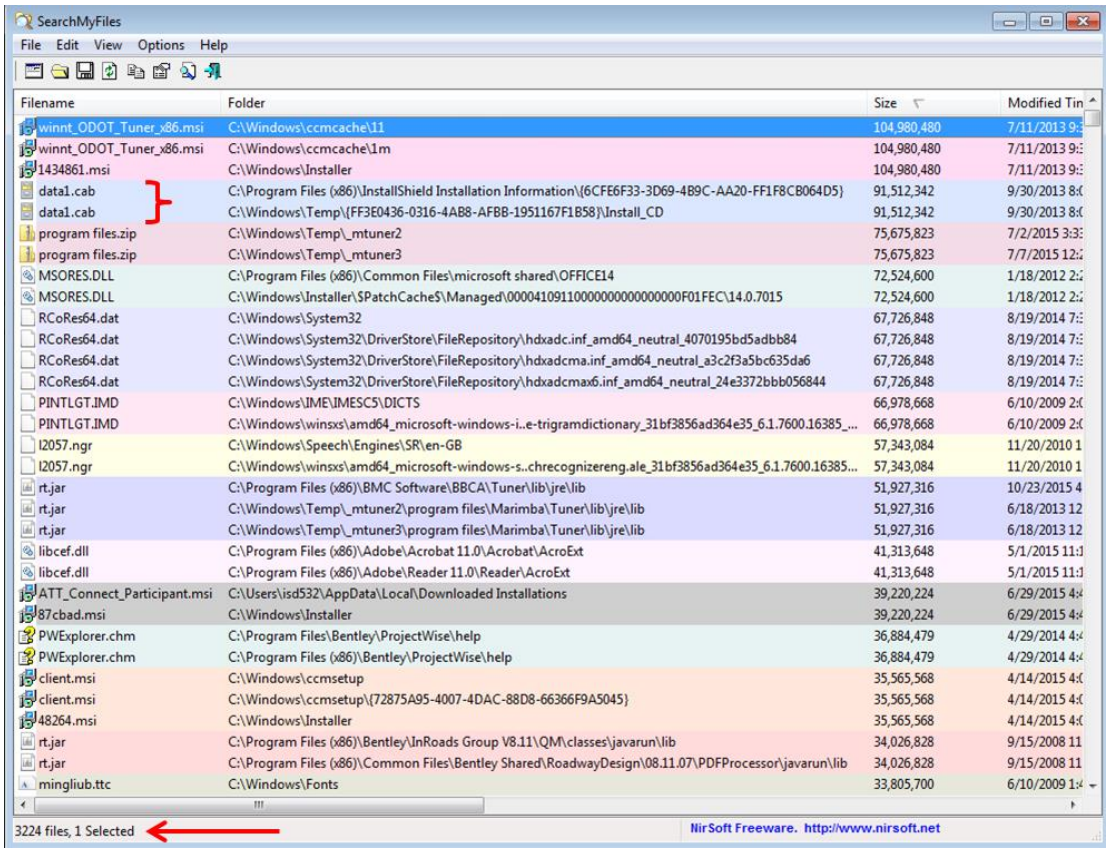



***Note:*** *If a large number of files are being scanned, the 'Checking Duplicates… %' in the lower-left corner may show 0. The percentage will increase as the process continues. Please allow sufficient time for the search to complete.*

*"Stop" is an option at the top of the* ***SearchMyFiles*** *main window until the search has completed. Once the search has completed, the "Stop" button is replaced with File, Edit, View, Options, and Help pull down menus.*



The colored shading of each group of duplicate files is intended as a visual aide. The total number of files found is shown in the lower-left corner of the **SearchMyFiles** main window indicating the search and processing is complete.




In the results window, using the right-click **Open Folder** command will open Windows Explorer to perform data management tasks.

## Saving the Results

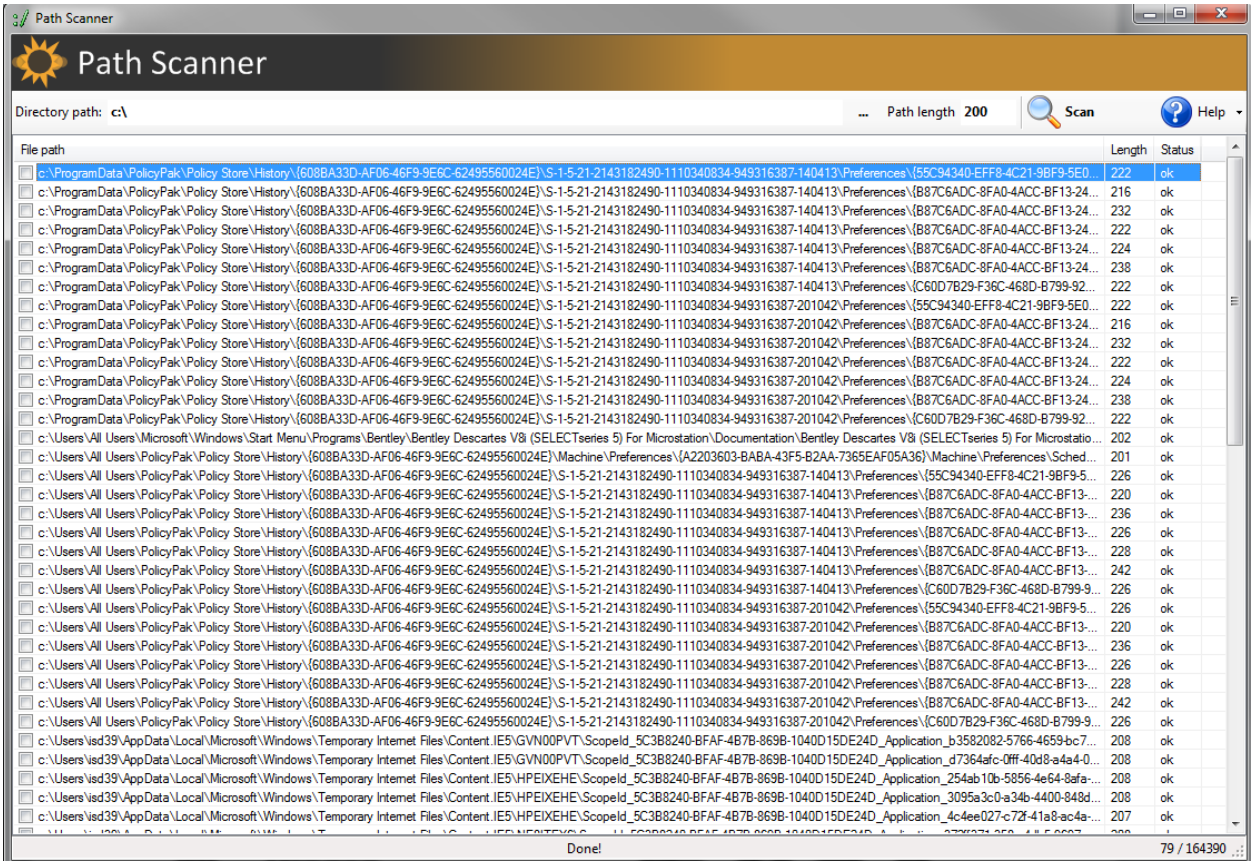
The results can now be used to perform data management tasks. It is NOT recommended to just start deleting files from this application. A good practice would be saving the results list into an Excel file for future reference and tracking of files that can be archived, renamed, or deleted.

To save the results, click on the title bar of the main window to activate it. Click **Edit > Select All** to select all results. Next click **File > Save Files Information**. This will open a save dialog. Navigate to a location to save the list of duplicate file ( C:\WORK is recommended). Change the Save as type to **HTML File – Horizontal (.htm)** which can import into Excel with preserved row shading and can also be viewed by a web brower. Enter a short descriptive file name like "Fdups" and click on Save to finish.

⚠ ***Warning:*** *It is* ***not*** *a recommended to use the SearchMyFiles Explorer edit functions. Use the File function Open Folder (F8) and Windows Explorer to manage duplicate files.*

# Path Scanner

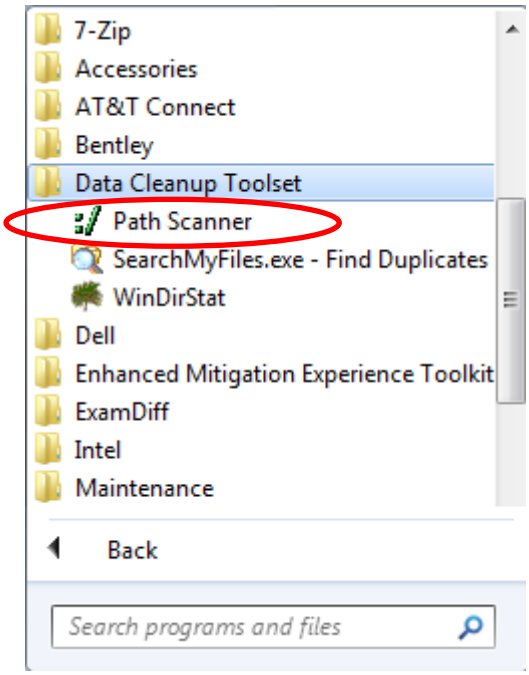
# Quick Start Guide



## Path Scanner - Find long paths

Path Scanner is a tool that scans a specified file path and shows you which paths are too long. You can specify a minimum length. Why is this important? Some Windows File Systems do not like file lengths larger than 255 characters. It can lead to errors when attempting to copy or move data. With Path Scanner you can avoid such errors by generating a list of paths that are longer than a certain length (ODOT will use a 200 character length limit). After the scan you can export the list to a CSV text file, which can be opened in Notepad or Notepad++ text editor, and use the text file to help in correcting the problem. This export function will be detailed further on in this guide.

## Launching Path Scanner

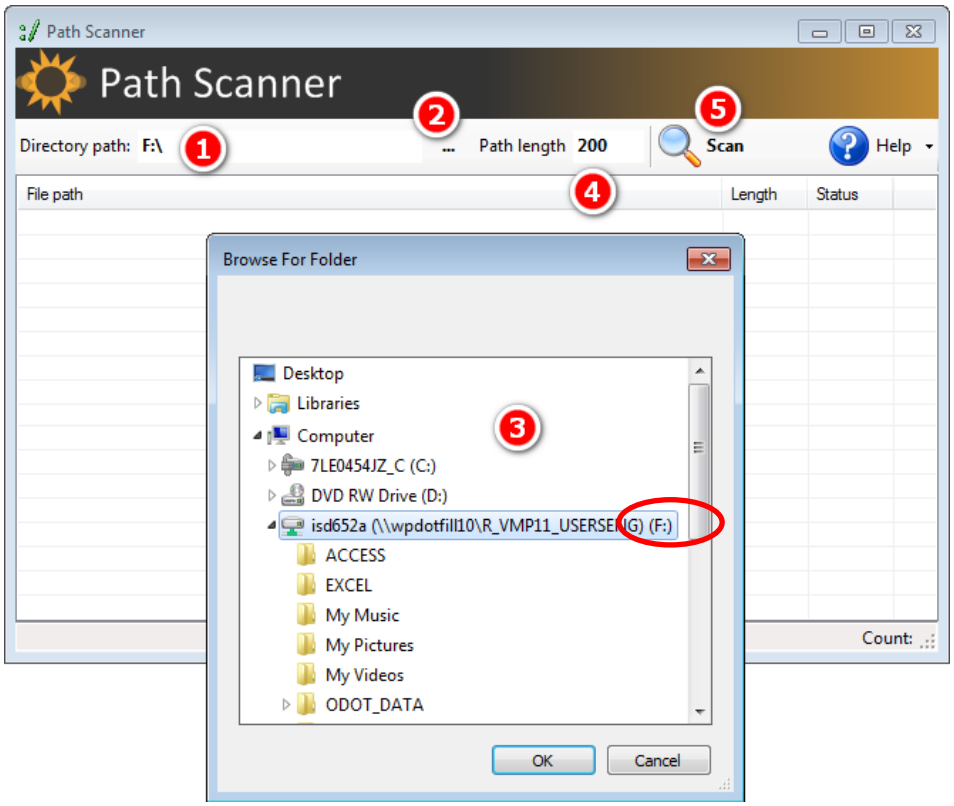



Once the application is installed, Path Scanner may be launched from the icon found in the Data Cleanup Toolset folder from the Start menu. Using the mouse, click on the Windows icon (Start button) , click on **All Programs**, scroll to and click on the Data Cleanup Toolset folder. You will see all three available tools. Click on the Path Scanner icon to launch the application.

# Using the Path Scanner

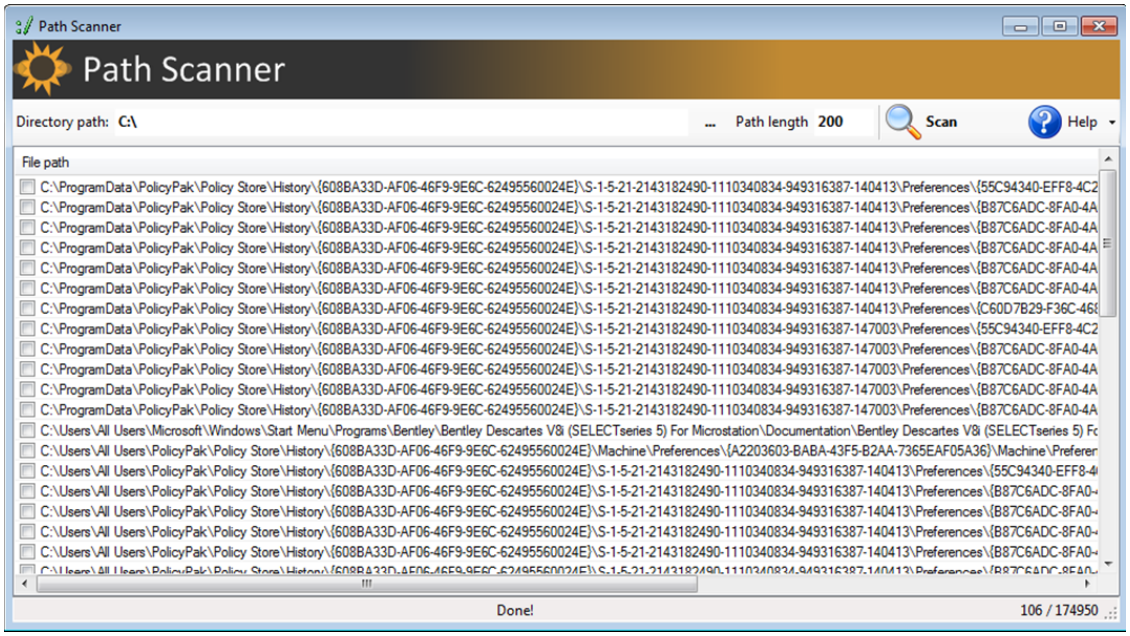
## Selecting the Drive to Scan

When the application is initially launched, you will need to change a few of the default settings. First, select a drive or folder path (**1**). You can manually type in the path, or you can click on the ellipsis (**…**) (**2**) to open a pop-up window to select options using the mouse. Typically, you will select the **F:** (personal network drive for ODOT users) (**3**). Next, change the default Path Length from the default value of 150 to **200** (**4**). Finally, click on the Magnifying Glass icon to start the scan (**5**).




## Reading the Results

When the scan has finished, you will see the results in the main window as seen in the picture below. In this example scan, the C drive was selected along with a 200 character path length. The results in the lower right corner indicate 106 paths out of 1744950 exceed 200 characters in length.

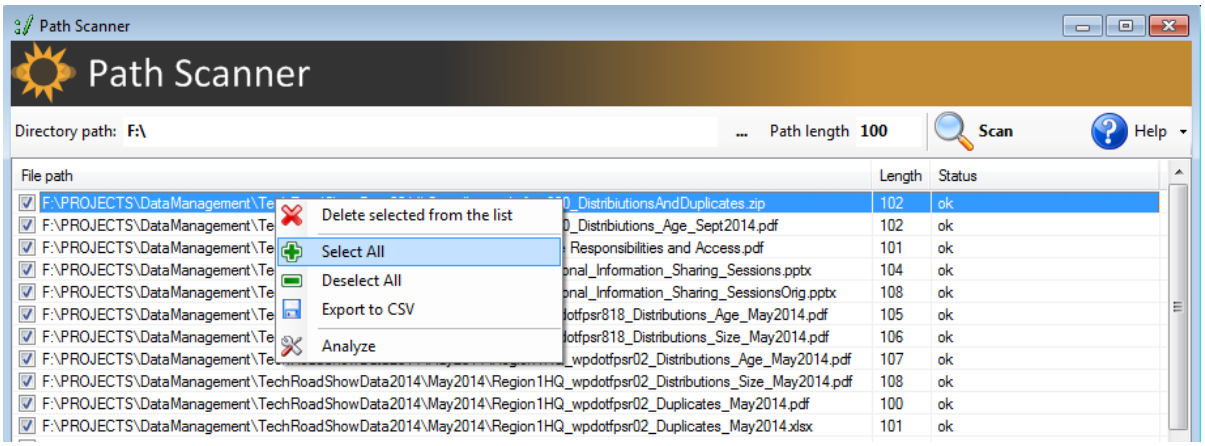



***Note:*** *A network path will not accurately reflect the 'actual' character count for the path length. This is why a path length of 200 is recommended to help account for network drives. This provides a buffer of 55 characters which remains within Microsoft character limitations and helps avoid other issues when working with data.*

*Example: ODOT engineering users have an F:\ drive where the actual path name is something like \\wpdotfill10\R_VMP13_USERSENG\HWYR00A. All these characters aren't included in the count.*

## Exporting and View the Results

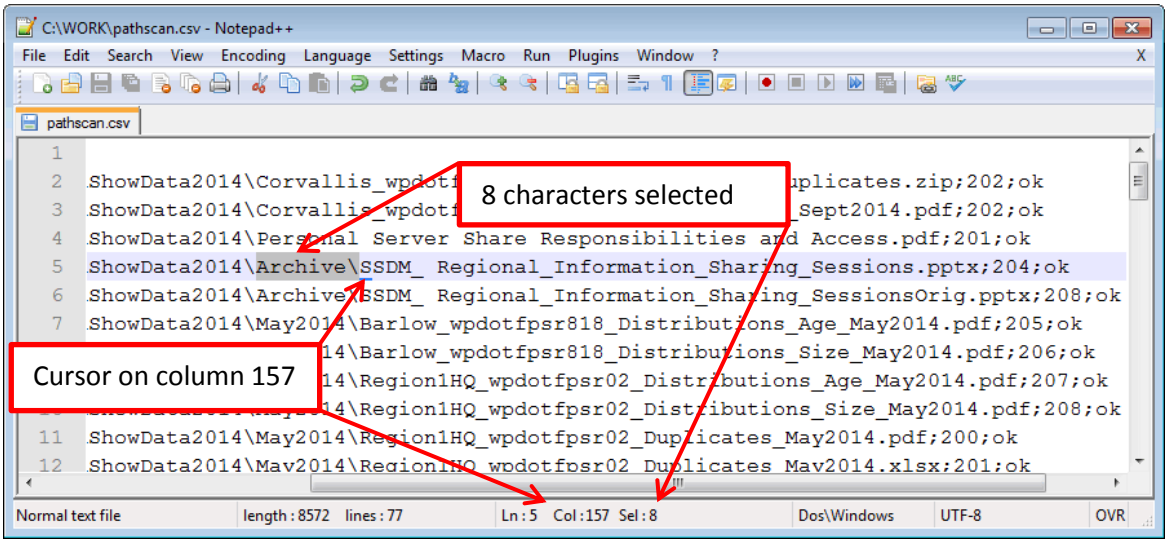
Once the scan is complete, you can right mouse click in the results window and choose one of several options. It is not recommended to use the Delete option from this application. The recommended action is to use the **Export to CSV** option, and a Text Editor to view the list. The example below used a path length of 100 for ease of illustration.




Right click on any result, and choose Select All. Note a check mark in the left column will indicate that that each row is selected. Right click again and choose the Export to CSV option. This action will open a Save As window to save it where you can find it like C:\WORK and a meaningful name like "pathscan.csv". This file may help serve as a working list to better track which files are being targeted for action steps.

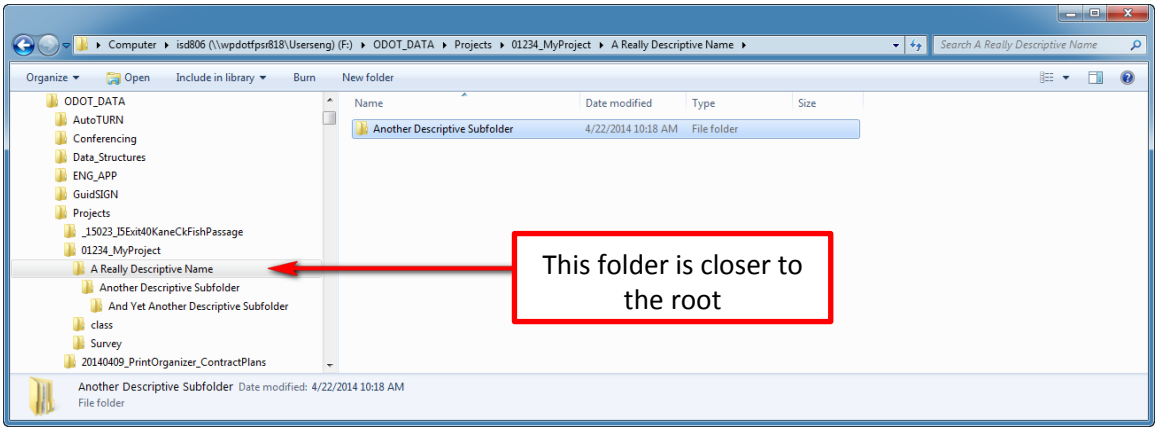
Open the file exported in a Text Editor like Notepad or Notepad++. The path, length and status are separated by semicolons.

Notepad++ is a free program that you can request to have installed. Notepad++ has nice features like showing the current column number and the number of selected characters. Like Microsoft Word, it maintains a list of recently opened files under the **File** menu.



## How to Shorten Paths

If your folder structures are way too deep and your filenames run on and on, your tree needs a good pruning. You will get the most out of the fewest edits if you shorten folder names closer to the root. First, pick a particular branch and exit any file that you may have open on that branch. Then, begin near the top (that would be the left-most side in Windows Explorer) and rename the folders so that they are shorter.




If you are storing STIP project data, consider using the key number and an abbreviated project name. For example, "54321 OR999 at Some Intersection or Creek" could become "54321OR999SomeCr". Notice the lack of spaces and using upper and lower case to make it more readable. I eliminated 25 characters!

If you've been saving email messages, you may want to shorten those long filenames. Try using numeric values to help your folders sort instead of using "A_", "AA_", or "AAA_" as prefixes to folder names. 1Zelda, 2Meeting, 3Admin, 4Project sort very nicely and your eye hardly sees the numeric prefix.

The above section was adapted from Long File or Folder Names are a Problem.

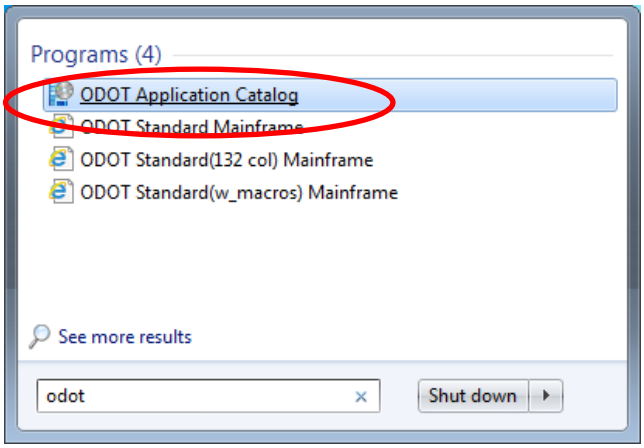
⚠ ***Warning:*** Do not rename folders that programs or other users are dependent on. MicroStation users have a folder F:\ODOT_DATA. If this folder is renamed then MicroStation will not run. You may need to check with your team before renaming a root folder in a crew share.

# Data Cleanup Toolset

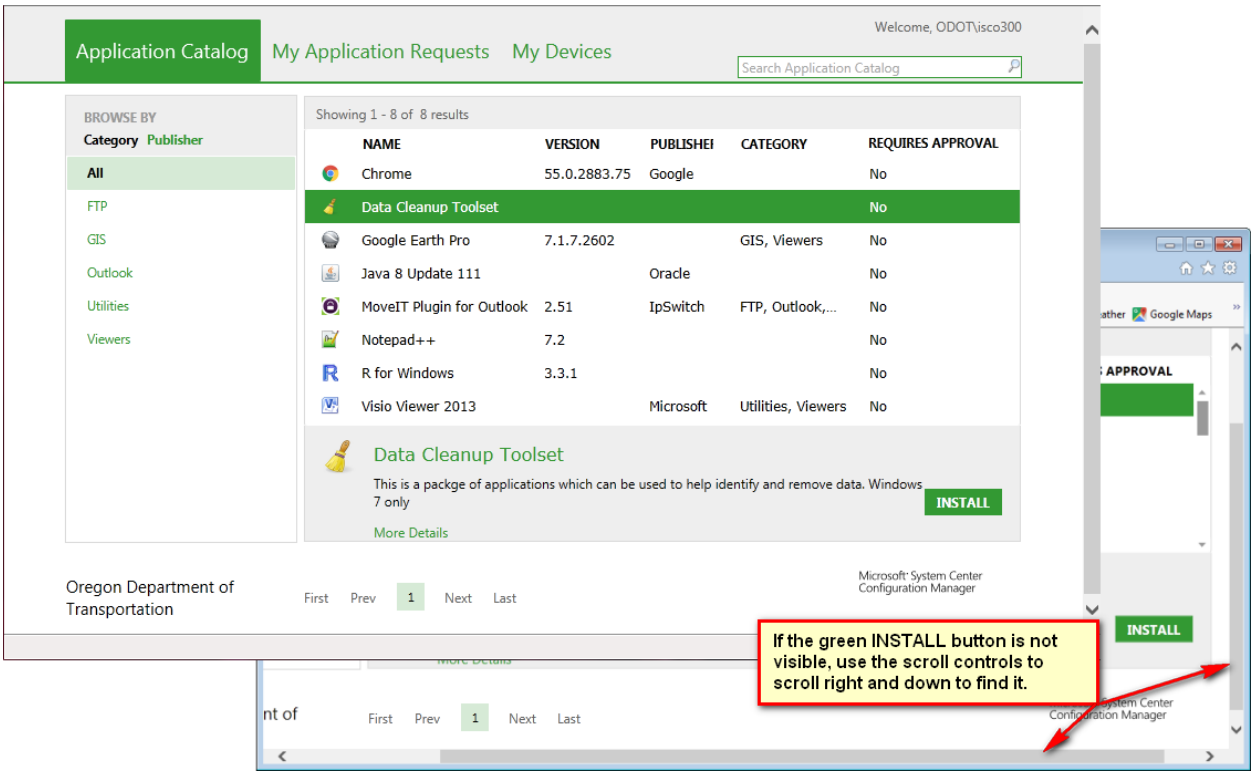
# Installation Guide

The Data Cleanup Toolset can be installed by anyone at ODOT by the following steps:

Close other applications because a restart may be required. Click the Windows Start Button and type **ODOT** into the Search Field. Select **ODOT Application Catalog** that appears at the top of the menu.

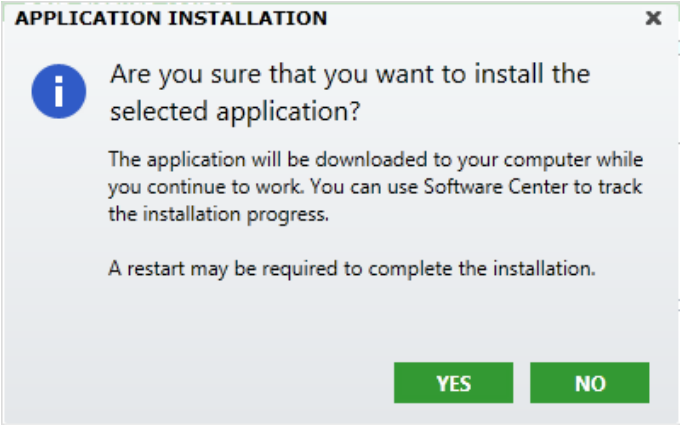



The ODOT Application Catalog will open in your web browser. Select **Data Cleanup Toolset**. If the list is long, you can select the Utilities category to narrow the list.

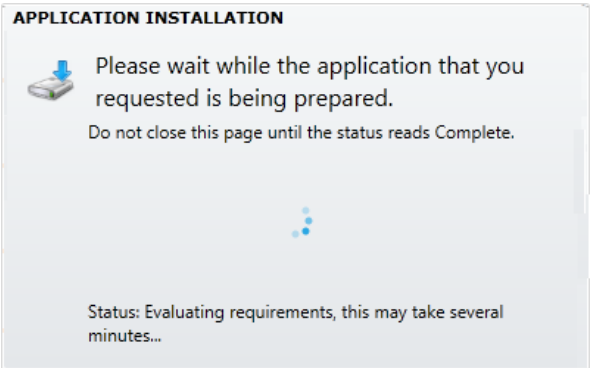



Click the green **Install** button to begin the installation.

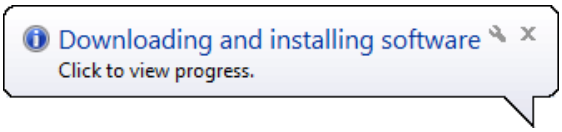
***Tip!*** *If you don't see the green Install button you may need to use the horizontal and/or vertical scroll bar controls to find the button.*



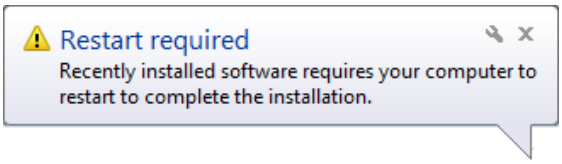

Click **Yes**. Note that a restart may be required.




Wait during the installation preparation.



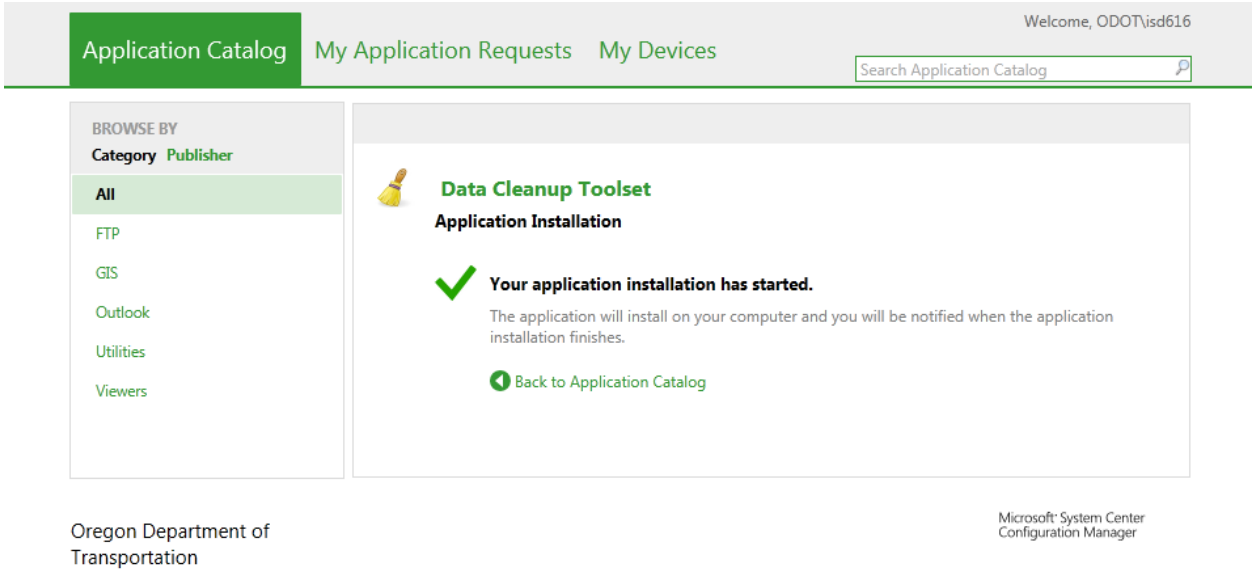

After a few minutes you may see notification on the task bar the software is downloading and installing.



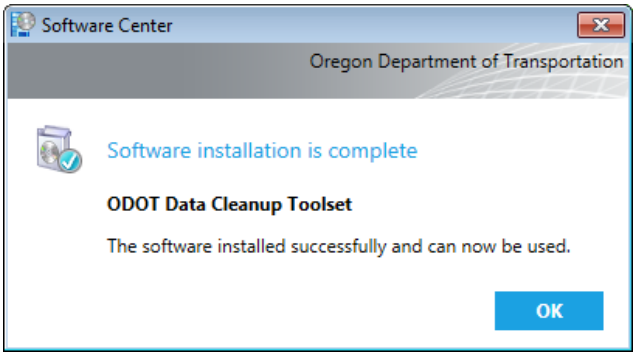

On older machines there are some dependencies that may be required to be installed first. These can take longer to install than the Data Cleanup Toolset and require a restart before the installation can continue. When prompted, restart the computer and the installation will continue after you log back on.

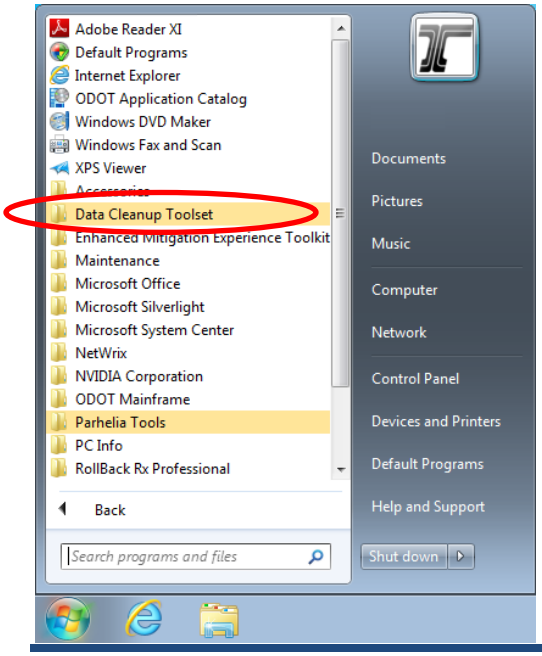
***Tip!*** *If you or your crew expect to install the Data Cleanup Toolset, the dependencies that take longer to install, can be pushed out to your computers ahead of time. Contact the Computer Support Desk or the TRC and they will have the dependencies installation added to your or your crew's computer maintenance.*

After the installation has started you will be notified in the Application Catalog.




Click the **OK** button when installation is completed. You may begin using any of the Data Cleanup toolset. This Final dialog box may appear under other windows or on a different screen.




There is no notification that the installation is done. A quick way to see if the installation is finished is to check the Windows Start Menu. On the Start Menu click **All Programs** and the list programs will appear. Look for **Data Cleanup Toolset** in the list of programs. Because it has just been installed, it should be highlighted as shown. You can now use the tools.