Gubbs Copy Data II 1.x.x

Copy Data II 1.x.x and earlier

User Manual

Gubbs, Inc.
265 Blue Spruce Circle
Alpharetta, GA 30005
P: 770-573-0169
F: 508-453-1338

gubbs@gubbsinc.com
www.gubbsinc.com

Version 2.0 (April 16, 2013)
G:\GubbsInc\WebSite\GubbsInc\GubbsInc03\files\CopyDataII_UserManual_002.doc

Table of Contents

Table of Figures

Table of Figures

1   Abstract

An age-old problem with scientific data acquisition is local workstation hard drive raw data administration. Local data acquisition workstation hard drives become full and require transfer, archival, backup or deletion of data acquisition raw data and related files. This is especially true for full-scan mass spectrometry workstations whose raw data files can be huge.

Manual data file backup/archive workflows (most often to a network shared drive) may be implemented, but can be troublesome and tedious:

- Manual copy/paste actions completed using Microsoft® Windows Explorer can be exceedingly slow, especially when trying to copy tens of gigabytes of data. There is a significant amount of overhead associated with the use of Windows Explorer.
- If an error occurs during a Microsoft Windows Explorer copy event, Windows Explorer terminates and it is unknown which data files have been copied and which have not been copied.

Many commercial-off-the-shelf (COTS) software application file copy tools are available and can be readily found in an internet search. Software solutions for file copy workflows because they use Windows API calls to copy/delete, which execute much more quickly than through Windows Explorer. However, few if any COTS file copy tools exist that can account for the nuances involved in scientific data file maintenance:

- It is often not advisable to immediately delete raw data files from the data acquisition workstation after a copy event in case it is required in the future to go back to the original raw data file (e.g. copy/paste event was corrupt or copied file was corrupted in the future for some reason).
- Some vendor data (e.g. Thermo® sequence files) contain sample-specific path columns that have referenced file paths and/or directory paths hardcoded in the column cell. If this type of data is moved and it is desired to continue to process the files, the path columns of the sequence file must be updated to the new directory paths and/or the referenced file paths after the data files have been moved.
- It is often desirable to maintain a certain directory structure in the new location that has meaning to the user. E.g. a data directory structure may be C:\[Vender]\[Data]\[Department]\[Year]\[Month]\. If would be desirable to have this data structure mimicked in the new directory location of the transferred data files.
- A scheduled data copy event should allow users to NOT copy/move data less than a certain number of days old. Without this type of feature, actively-acquiring or required data files will be copied/moved which may interfere with the operation of the data acquisition system (e.g. an actively-acquired Sciex® Analyst® .wiff file or scheduled Thermo® QuickQuan® daughter directory sequence files).
- Some data files (e.g. template instrument method files) should never be moved because they may be needed in the normal operation of the data acquisition instrument.

- Some levels of data directory structure may need to be maintained in order to aid in data acquisition. E.g. All the data and empty directories in the underlined portion of this data directory may be allowed to be deleted, but the non-underlined portion may be required to remain for future data acquisition events:

  C:\Xcalibur\QuickQuan\Data\Microsomal Metabolic Stability\2010\01-29\RLM\01252010_145128\01_MS012510_1\

2   Proposed Workflow

From a high level, the following would be a desirable workflow for an automated data acquisition workstation backup software solution (either manually executed within the software or ran using a scheduler within the software).

1. COPY files from one or more user-defined directories on the local data acquisition workstation to the network.

   a. A defined portion of the data file path will be preserved in the network location.
   b. Ignore data files that are older than a user-defined number of days.

      This will prohibit the software from attempting to copy/move actively-acquiring raw data files (e.g. .wiff or .sld).

   c. Ignore files located in user-defined Excluded Directories.

      E.g. Data acquisition system required template files, etc.

   d. Automatically modify any Thermo sequence file (.sld) path columns to reflect the new location of the data.

2. MOVE the just-copied files from the data acquisition workstation to a temporary location (preferably on the local data acquisition workstation).

   a. This action cleans up the data acquisition directories, but the raw data files have been moved, not deleted.
   b. Ignore files that are older than a user-defined number of days.
   c. Ignore files located in user-defined Excluded Directories.
   d. Ignore files with user-defined file extensions.
   e. Delete the production directory(s) files and empty directories (except for Excluded Directories) after a certain level of directory structure.

3. DELETE the files located in the temporary location that are older than a certain number of user-defined days.

   a. The user-defined number of days can be set at such a number that it is highly probable that there would be no need to access the original raw data file.

3 Gubbs® Inc. Copy Data II software solution

Gubbs Inc. Copy Data II has all the features described in the proposed workflow. As well, it contains the following additional features:

- Accepts command line switches that can be used to call Copy Data II from an external application.

    o Different copy scenarios may be stored in different .ini files and called from different desktop shortcuts
    o Copy Data II may be called with different command line switches to perform specific instructions

- Log files are created detailing each copy/move/delete/.sld modification action.

    o For regulated environments, logs may be saved as .pdf or .xps.

- Exclude files with user-defined file extensions or user-defined locations.
- Has raw file vendor-specific features

    o Thermo

        - Automatically update sequence file (.sld) Path, Processing Method, Calibration Method, and Instrument Method columns.
        - Ignore copying Thermo date-time-stamped sequence audit files.

4 Copy Data II Operation

Below is a figure showing the Copy Data II main window. A description the general operation of Copy Data II follows.
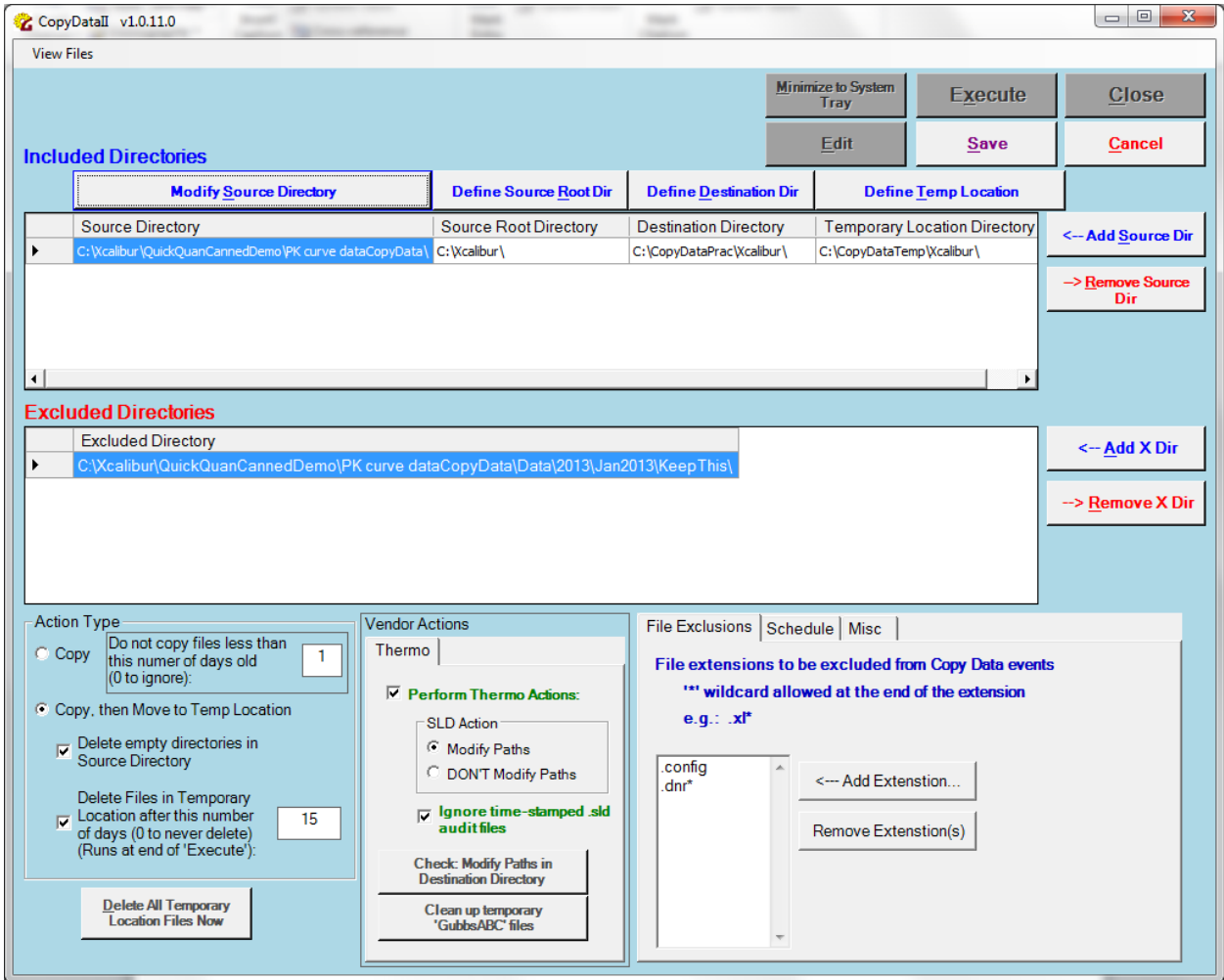

Figure 1   Copy Data II window – initial view

4.1   Included Directories Table

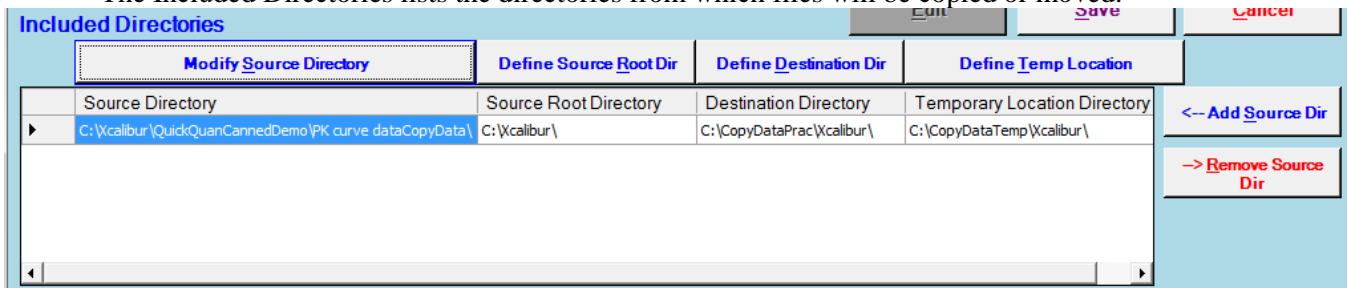The Included Directories lists the directories from which files will be copied or moved.

Figure 2   Included Directories

4.1.1   Define Source Directory

Users define one or more Source Directories (directories from which files are to be copied/moved) by clicking the blue 'Add Source Dir' button located to the right of the Included Directories table.

- Once configured, the Source Directory may be modified by selecting a row and clicking the blue 'Modify Source Directory' button located above the Source Directory column.

Users may remove one or more Source Directories by selecting the directory(s) and clicking the red 'Remove Source Dir' button located to the right of the Included Directories table.

4.1.2   Define Source Root Directory

Users must define the Source Root Directory by selecting a row and clicking the blue 'Define Source Root Dir' button located above the Source Root Directory column. The 'Root Directory' is the portion of the directory path that will be replaced by the Destination Directory when files are copied.

When the button is clicked, the user is presented with the window shown in Figure 3 below:
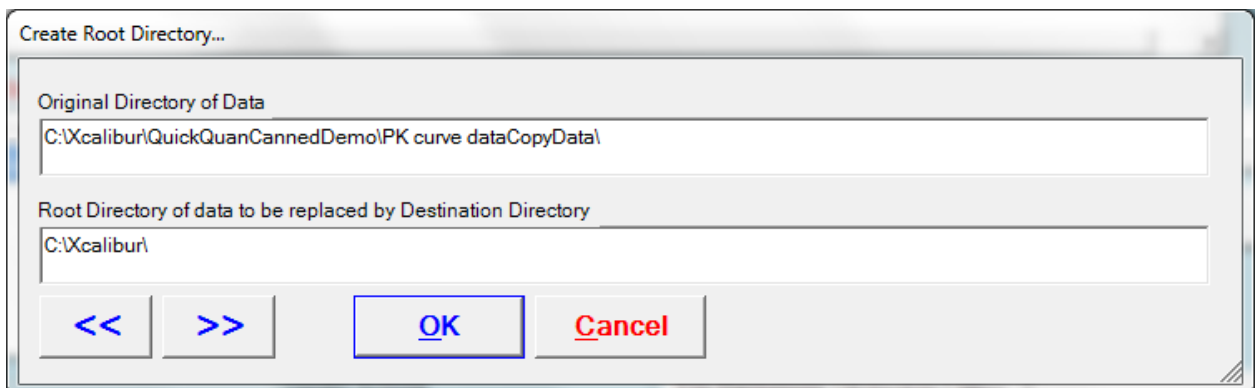
Figure 3   Create Root Directory window

> In the example above, the user has chosen that the 'C:\Xcalibur\' portion of the Original Directory is to be replaced by the path defined in the Destination Directory portion of the Included Directories table (see Section 4.1.3)
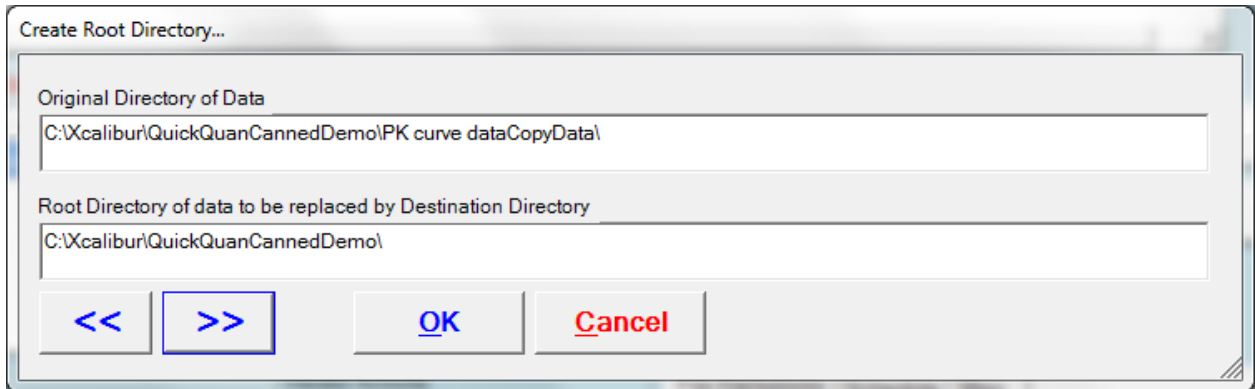

Figure 4   Create Root Directory window – modified choice

> If it is desired that the portion of the 'C:\Xcalibur\QuickQuanCannedDemo\' portion of the Original Directory is to be replaced by the path defined in the Destination Directory, then the user would click the blue '>>' arrow (see Figure 4 above):

4.1.3   Define Destination Directory

Users must define the Destination Directory by selecting a row and clicking the blue 'Define Destination Dir' button located above the Destination Directory column. The 'Destination Directory' is the root directory to which files will be copied or moved.

For example in the configuration shown in Figure 1, files located in the Source Directory (C:\Xcalibur\QuickQuanCannedDemo\PK curvedataCopyData\) will be copied to the Destination Directory (C:\CopyDataPrac\Xcalibur\). Since the Root Directory is defined as "C:\Xcalibur\", the new path of files will be "C:\CopyDataPrac\Xcalibur\QuickQuanCannedDemo\PK curvedataCopyData\".

4.1.4   Define Temp Location

If Action Type is set to 'Copy, then Move to Temp Location', then users must define the Temporary Location by clicking the 'Define Temp Location' button located above the Temporary Location Column.

4.2 Excluded Directories Table

The Excluded Directories Table lists the directories that will be ignored when copying files
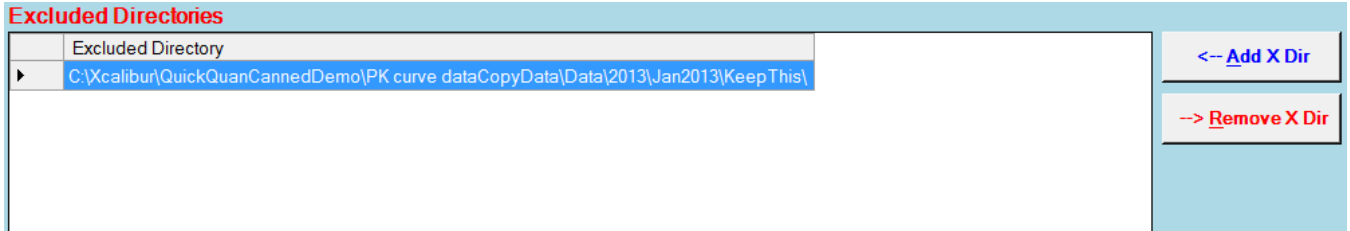

Figure 5 Excluded Directories Table

4.2.1 Add X Dir

Click the 'Add X Dir' button to add directories to be excluded.

4.2.2 Remove X Dir

To remove directories from this table, select one or more rows and click the 'Remove X Dir' button.

4.3 Action Type

There are two types of Copy Actions that can be performed by Copy Data II:
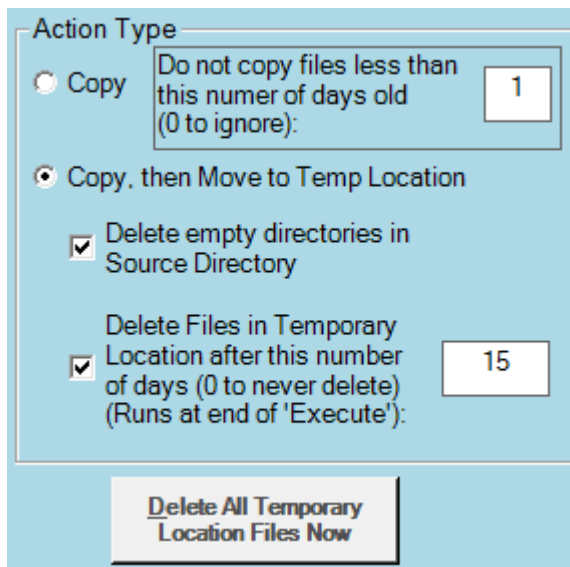
o Copy
o Copy, then Move to Temp Location

Figure 6 Action Types

4.3.1   'Do not copy files…' box

Enter '0' to ignore this feature.

This box applies to both 'Copy' and 'Copy, then Move…' action types. If Copy Data II is being run on a schedule and the Action Type is 'Copy, then Move…', it is possible that the event may occur during an active data acquisition event on the data acquisition workstation. It would not be desirable to Move files during an active data acquisition event because movement of files may cause the data acquisition system to fail.

Users can tell Copy Data II to ignore files that are older than, say, 1 day old, which would most likely ignore files included in an active data acquisition event. Of course, if typical data acquisition includes sequences lasting longer than 2 days, then the number should be set to '2'.

When performing a manual Copy event (clicking Execute button), the user is given the option to override the 'Do not cope files…' setting and copy or move everything.

4.3.2   'Copy' Action Type

If 'Copy' is chosen, then files will be copied to the Destination Directory and WILL NOT be moved from the Source Directory.

4.3.3   'Copy then Move to Temp Location' Action Type

If 'Copy then Move to Temp Location' is chosen, files will be copied to the Destination Directory, and then MOVED to the Temporary Location defined in the Included Directories table.

- o   Files located in Excluded Directories will be ignored.

4.3.3.1   'Delete empty directories in Source Directory' checkbox

If checked, Copy Data II will attempt to delete empty directories located in the Source Directory.

To maintain a certain level of directory/subdirectory structure, do this:

- ▪ Place a text file named 'KeepDir.dnr' (or some file extension of your choice) in the directory whose structure you wish to maintain.
- ▪ In the File Exclusions tab of Copy Data II, add the file extension ".dnr".
- ▪ Since the directory will have a file after Copy/Move events have completed, the directory/subdirectory will not be deleted.

4.3.3.2   'Delete Files in Temporary Location…' checkbox

Enter '0' to ignore this feature. Entering '0' means that files located in the Temporary Location will never be deleted automatically by Copy Data II. Files can be deleted manually in Windows Explorer or by clicking the Copy Data II – 'Delete All Temporary Location Files Now' button.

Files located in the Temporary Location that are older than the entered number of days will be deleted. The deletion event occurs at the end of any Copy event.

4.3.4   'Delete All Temporary Location Files' button

Click this button to manually clear all files located in the Temporary Location Directory.

4.4   Vendor Actions

Copy Data II can perform actions that are unique to specific vendor files.
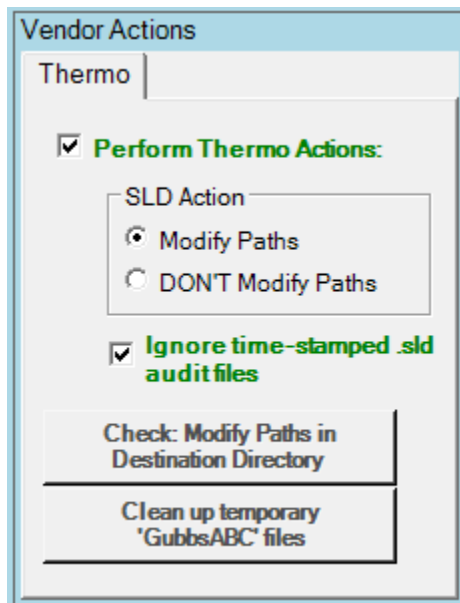
4.4.1   Thermo

Figure 7   Thermo Vendor Actions

A desirable action for users is to copy all acquired data to a network location and process data there. This allows the user to take advantage of company-level backup procedures to automatically backup and/or archive processed data.

A problem with Thermo data is that data is processed according to acquired (.raw) and processing method (.pmd), instrument method (.meth), and calibration (.xcal) files whose locations (paths) are defined in the sequence (.sld) *at the time of acquisition*. If all required data files are moved to a new location, users must user Xcalibur® to manually open the .sld file and re-define the paths of the .raw, .pmd, .meth, and .xcal files.

Copy Data II can, if desired, automatically open the .sld and modify the .raw, .pmd, .meth, and .xcal files to reflect their new location.

4.4.1.1 'Perform Thermo Actions' checkbox

Users check this box to activate Thermo specific features in Copy Data II

4.4.1.2 'SLD Action' option group box

4.4.1.2.1 'Modify Paths' option button

Users select this button if they want Copy Data II to modify the paths of the .raw, .pmd, .meth, and .xcal files to reflect their new location after a copy event.

4.4.1.2.2 'DON'T Modify Paths' option button

Users select this button if they DON'T want Copy Data II to modify the paths of the .raw, .pmd, .meth, and .xcal files to reflect their new location after a copy event.

4.4.1.3 'Ignore time-stamped .sld audit files' check box

Xcalibur (and Xcalibur controllers such as Aria/Transcend) sometimes create .sld copies that are date-time stamped and are meant to be audit trails of .sld events. Sometimes large numbers of these audit files are created and clutter the data acquisition directory.

If users check this box, Copy Data II will ignore .sld audit files and not include them in the Copy event and will not modify their paths.

4.4.1.4 'Check: Modify Paths in Destination Directory' button

The very first time Copy Data II is run manually on a Source Directory, the entire process can be long if the Source Directory has a large amount of data (e.g. tens to hundreds of gigabytes) and the user may have to kill the process in order to free up the instrument to acquire data. Killing the process can leave some .sld files in a not-modified state.

Users can click this button to inspect all the .sld files located in the Destination Directory(s) to check if the .sld paths have been updated and modify if needed.

4.4.1.5 'Clean up temporary 'GubbsABC' files' button

In the .sld modification process, Copy Data II must make a temporary file whose file name contains the fragment 'GubbsABC'. These files are deleted at the end of the process. In the scenario describe above in Section 4.4.1.4, these temporary files may not get deleted.

Users can click this button to delete all instances of 'GubbsABC' .sld files.

4.5   File Exclusions Tab

Files with certain extensions may be configured to be ignored in the Copy or Move event.

Note that wildcards may be used (e.g. use .xl* to account for .xls, xlsx, .xlsm, etc.)
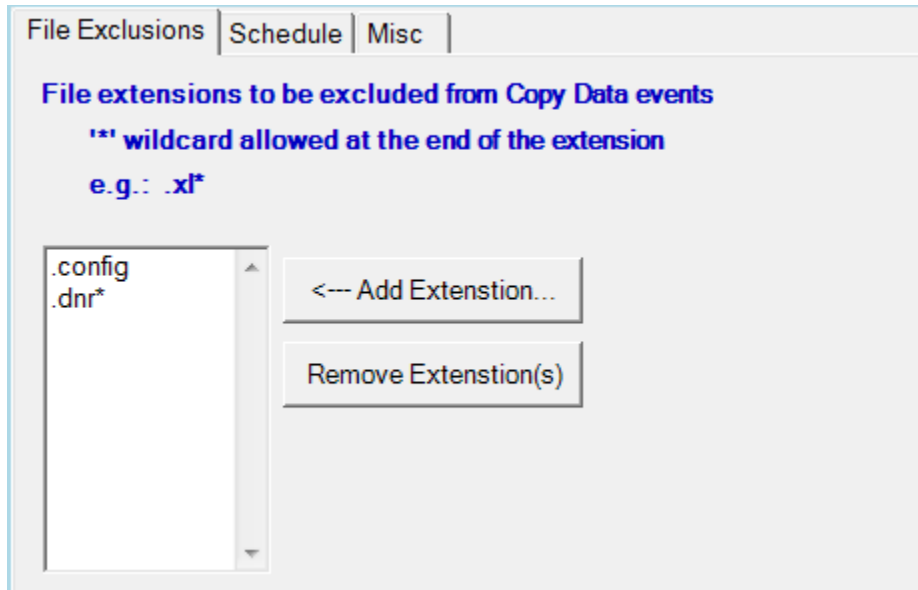

Figure 8   File Exclusions Tab

4.5.1   'Add Extension' button

Users can click this button to add file extensions to be ignored in the Copy or Move event.

4.5.2   'Remove Extension(s)' button

To remove extensions from the list, select one or more items and click this button.

4.6   Schedule Tab

Copy Data II can be minimized to the System Tray and ran automatically according to a schedule.
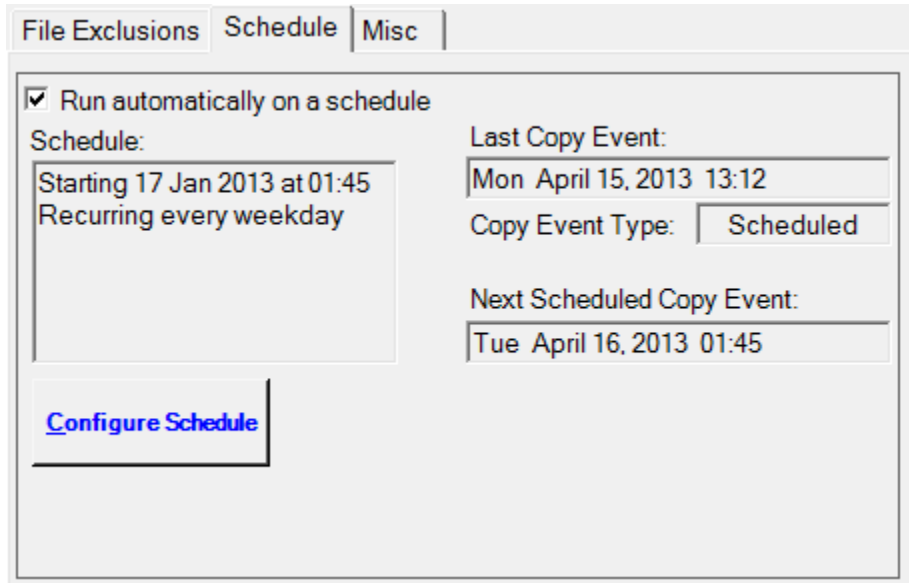

Figure 9   Schedule Tab

4.6.1   'Run automatically on a schedule' check box

Check this box if users wish to run Copy Data II on a schedule.

Note: In order to minimize to the system tray after a scheduled copy event, users must check the 'Minimize to System Tray' check box located in the Misc Tab of Copy Data II (see Section 4.7.2).

4.6.2   'Schedule' text box

This text box describes the currently defined schedule. If the 'Run automatically…' check box is unchecked, this text box will be empty or contain 'None'.

4.6.3   'Last Copy Event' text box

This text box lists the last time any Copy event occurred.

4.6.4   'Copy Event Type' text box

This text box lists the type of Copy event that last occurred (e.g. Scheduled or Manual).

4.6.5   'Next Scheduled Copy Event' text box

This text box lists the next scheduled Copy event.

4.6.6   'Configure Schedule' button

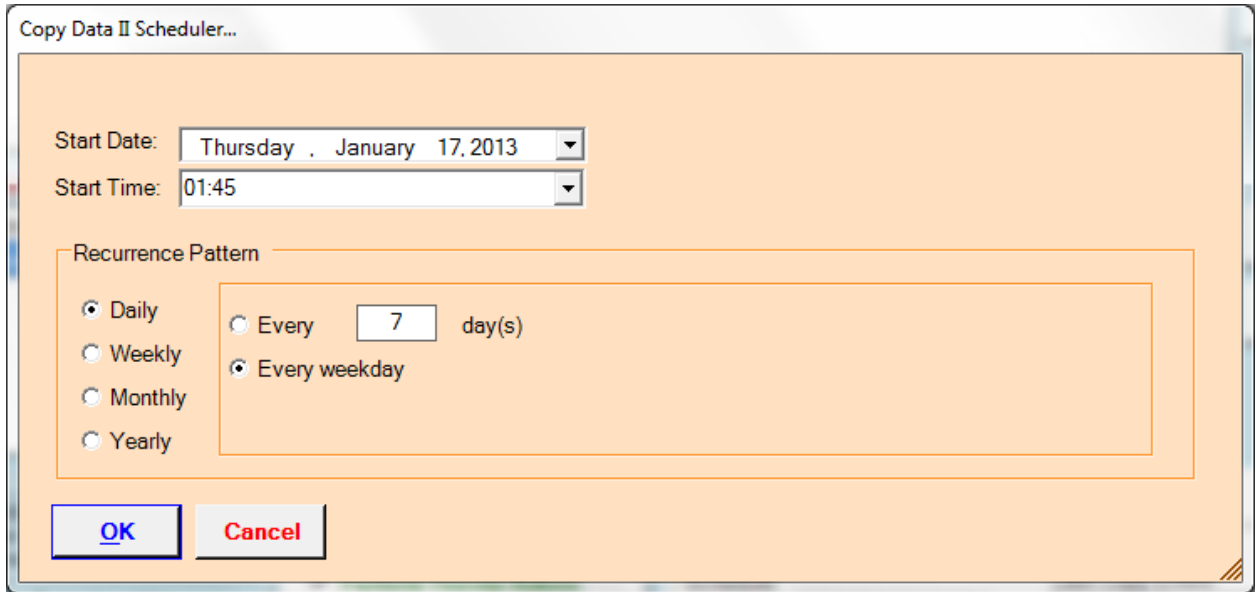Users click this button to configure a schedule (see Figure 10 below).

Figure 10   Copy Data II Scheduler

4.7   Misc Tab

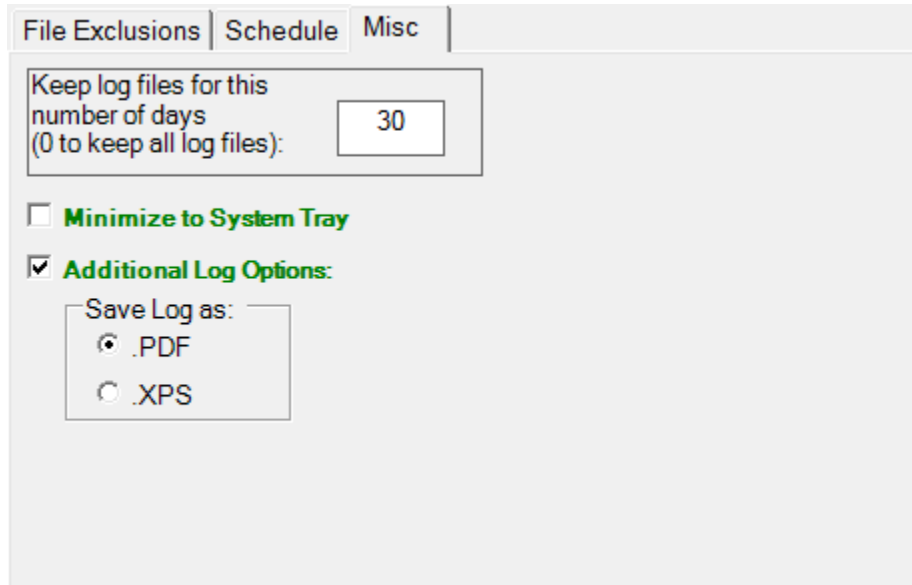Copy Data II has several miscellaneous settings.

Figure 11   Misc Tab

4.7.1   'Keep log files…' text box

Enter '0' here to keep all log files

If a number other than 0 is entered, log files whose age is greater than the number of days entered will be deleted at the end of a Copy event.

4.7.2   'Minimize to System Tray' check box

If checked, Copy Data II will be automatically minimized to the System Tray after a Copy event.

4.7.3   'Additional Log Options' check box

Log files are saved as .txt files. If this box is checked, users can additionally have the log files saved as Portable Document Format (.pdf) or XML Paper Specification (.xps). These file types are un-editable files.
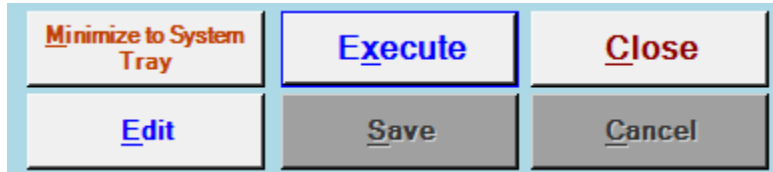
4.8   Top Level Buttons

Figure 12   Top Level Buttons

4.8.1   'Minimize to System Tray' button

Users click this button to manually minimize Copy Data II to the system tray.

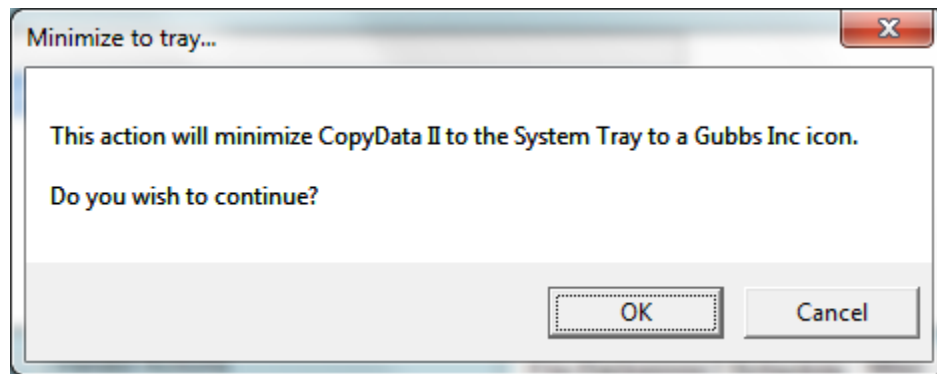Users will be prompted with the following dialog box:

Figure 13   Minimize to System Tray dialog box

4.8.2   'Execute' button

Users click this button to manually execute a Copy event
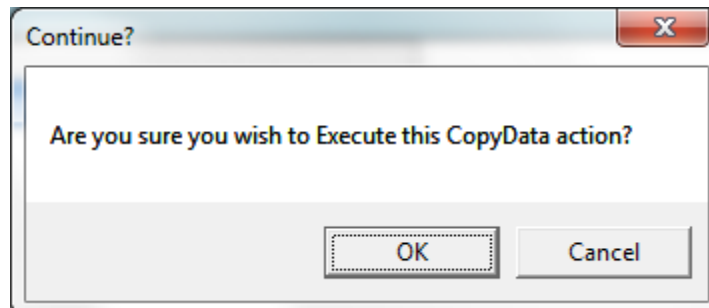
Users will be prompted with the following dialog boxes:

Figure 14   Execute dialog box 1

If the 'Do not copy files less than…' text box contains a value other than 0, the user will be asked if this setting is to be overridden (i.e. copy or move all files, no matter their age):
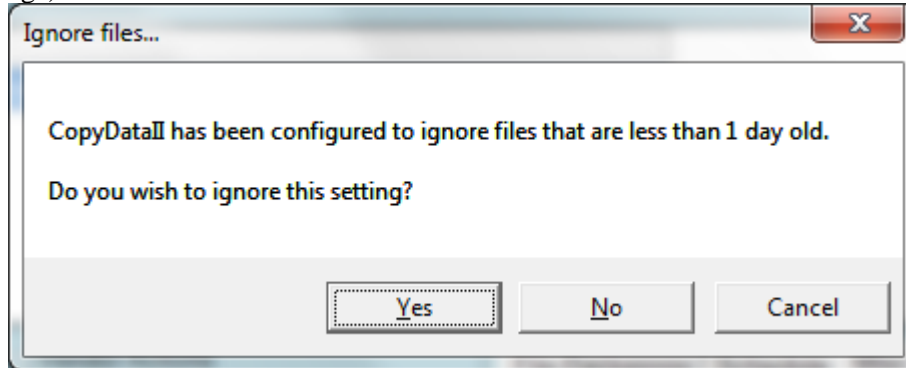

Figure 15   Execute dialog box 2

4.8.3 'Close' button

Users click this button to close Copy Data II.

If the 'Minimize to System Tray' check box (see Section 4.7.2) is checked, users will be prompted with this dialog box:
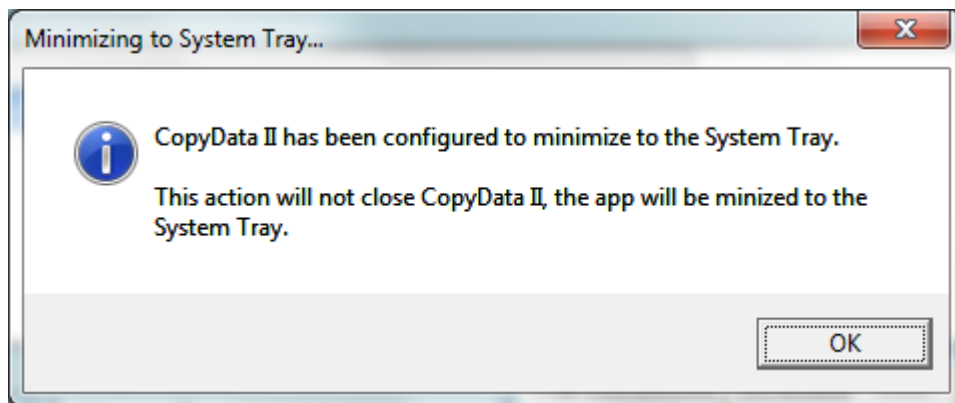

Figure 16   Close dialog box

4.8.4 'Edit' button

Users click this button to put Copy Data II in the edit mode to make changes to settings.

4.8.5 'Save' button

Users click this button to save any changes made to Copy Data II settings.

4.8.6 'Cancel' button

Users click this button if users make changes to Copy Data II settings while in edit mode and wish to discard all changes.

5   Command Line Switches

Copy Data II has several switches that allow users to call Copy Data II from an external program (e.g. the command line or from a Copy Data II shortcut). Note that any combination of switches may be included in the argument string passed to Copy Data II

5.1   '.ini' switch

Copy Data II settings are saved in a .ini file named "C:\GubbsInc\CopyDataII\Ini\CopyDataII.ini". Users can create different settings and save them as different .ini files and call these different .ini files in a command prompt.

For example, a shortcut to the CopyDataII.exe file may be placed on the desktop with the following argument string in the Target box of the shortcut Properties Window:

C:\Program Files\Gubbs Inc\CopyDataII\CopyDataII.exe
"C:\GubbsInc\CopyDataII\Ini\CopyDataII_Long.ini"

When the shortcut is run, the settings from the .ini file "CopyDataII_Long.ini" will be loaded.

5.2   DoAuto switch

If CopyDataII.exe is called from the command line with the switch DoAuto contained in the argument string, Copy Data II will:

   o   Open
   o   Immediately execute a Copy event
   o   Close (or minimize to the system tray if so configured)

For example: C:\Program Files\Gubbs Inc\CopyDataII\CopyDataII.exe DoAuto

5.3   /S switch

Users can pass a Source Directory to Copy Data II that will override the Source Directory(s) defined in Copy Data II.

For example: C:\Program Files\Gubbs Inc\CopyDataII\CopyDataII.exe DoAuto \S "C:\Xcalibur\00021\"

If this argument string is passed, Copy Data II will:

   o   Open
   o   Immediately execute a Copy event
       o   The Source Directory(s) configured in Copy Data II will be overridden with "C:\Xcalibur\00021\"
   o   Close (or minimize to the system tray if so configured)

5.4   .sld switch

Thermo Xcalibur has a feature that allows users to call an external application after the execution of a processing method .pmd. This feature has a switch (%S) that will pass the path of the sequence file .sld to the external application

Copy Data II can interpret the .sld path passed to it and use it to override the currently configured Source Directory(s).

The following is a suggested workflow to automatically execute Copy Data II using the sequence path passed to Copy Data II from the Xcalibur .pmd.

1. The user would configure a special .pmd that has enabled a Run Program action
   a. Activate the Programs window from the menu View – Programs
   b. Check the 'Enable' box in the first column so that the value changes to 'Yes'
   c. Action is 'Run Program'
   d. The 'Program or Macro Name' is the path to the installed Copy Data II.exe
   e. Parameter DoAuto tells Copy Data II to execute and close
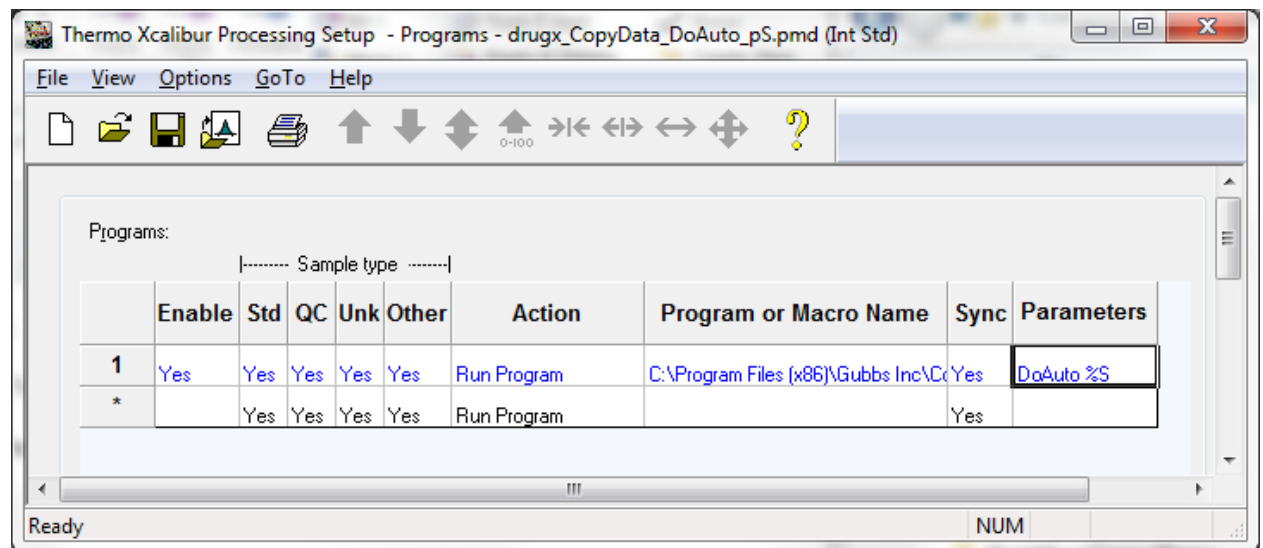   f. Parameter %S passes the path of the .sld to Copy Data II to override the Copy Data II Source Directory(s)

Figure 17   Thermo Xcalibur Processing Setup

2.   The user would then configure an additional blank sample at the end of the sequence that contains the Copy Data pmd.
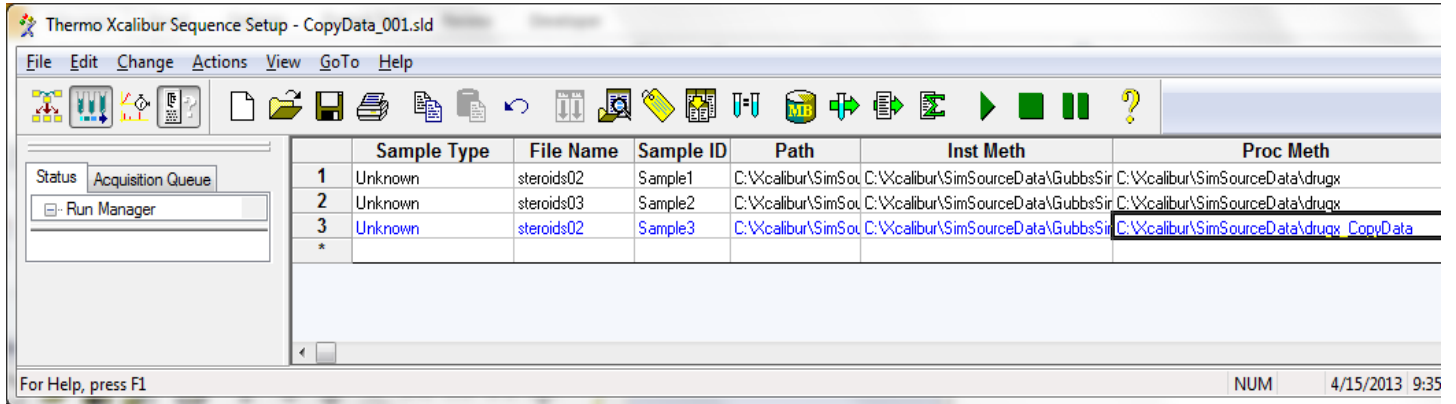

Figure 18   Thermo Xcalibur Sequence Setup

3.   When starting acquisition, the user would check Programs in the Processing Actions section (Note: Quan does not need to be checked).
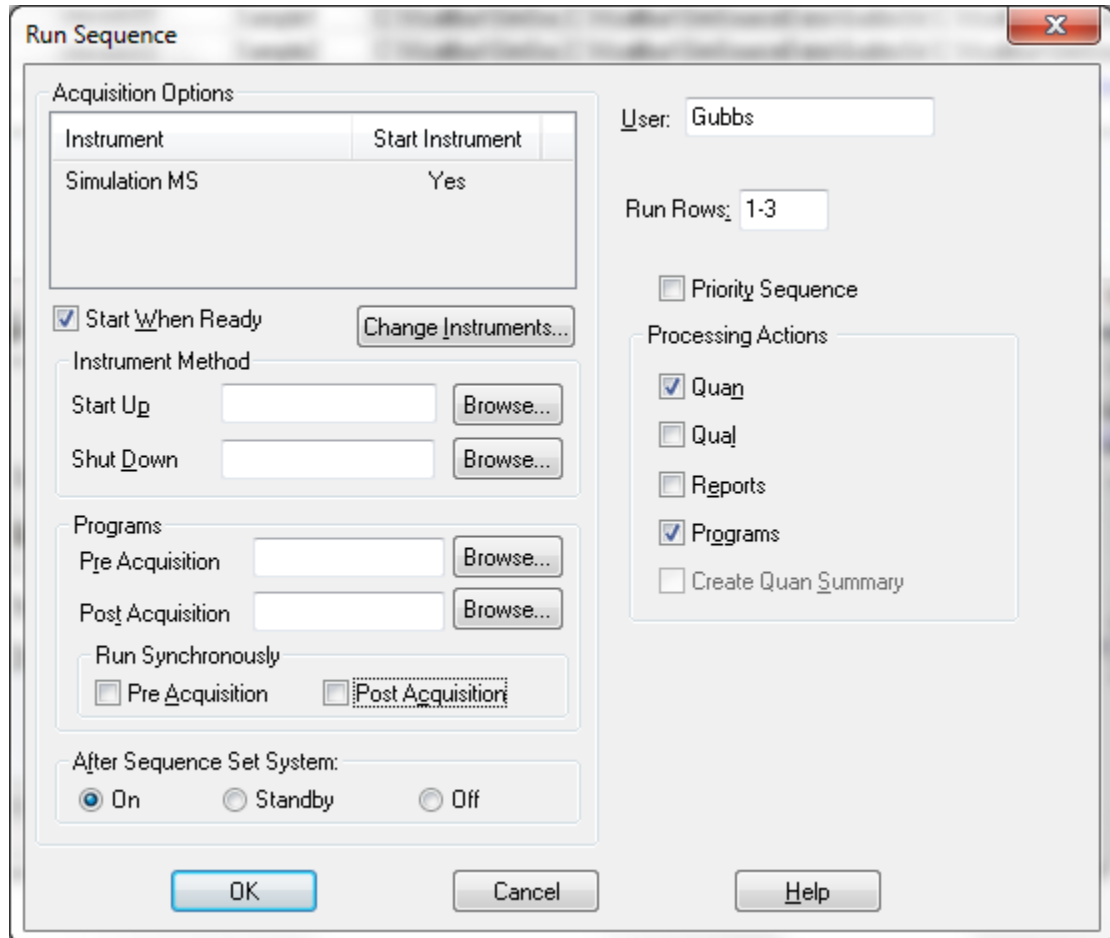

Figure 19   Thermo Xcalibur Run Sequence

4.   Copy Data II will open, execute, and close. All this is done in the background transparent to the user.

6   System Specifications

Item	Requirement
Operating System	Windows 7 (SP0 or SP1), 32-bit or 64-bit
DotNet	DotNet 4.0 (supplied in Copy Data II installer)
If Thermo Actions	Xcalibur 2.0.7 or newer (Thermo XDK MUST be installed
Drive Space	1 Gb
RAM	250 Mb