

Auditing Electronic Data

Tips and Techniques for Utilizing Excel® to Audit Electronic Data



Rick Hoffman
Monsanto Company

January 30, 2014

Agenda

- ❑ Define “electronic data”
- ❑ Requirements for electronic data
- ❑ Tips for auditing e-data using Excel
- ❑ Watch outs



What Is Electronic Data?

- ❑ *FDA 21 CFR Part 11 - Electronic record* means any combination of text, graphics, data, audio, pictorial, or other information representation in digital form that is created, modified, maintained, archived, retrieved, or distributed by a computer system.
- ❑ PC magazine - Refers to data of any type stored as binary code in the computer, a hard disk or other computer storage device, or to any data transmitted in binary code.
- ❑ Today's working definition – Alphanumeric data stored in a human readable format in an electronic system.



Direct Input vs. Transcribed Data

❑ Direct Input

- Data is typed (or captured) directly into an electronic device.

❑ Transcribed Data

- Data is recorded manually and then entered into the electronic device after the fact.

❑ Migrated Data

- Data is copied from one electronic format/location to another.



Basic Requirements – Direct Input

- Must be in a human readable format
- Must include these minimum elements for a unique record:
 - Study ID
 - Observer (and scribe)
 - Collection date/time
 - Variable ID
 - Units
 - Treatment ID
 - Value
- Additional information that may be required to uniquely identify an observation (e.g. plant ID, replicate, location)
- Ideally a validated system with an audit trail



Basic Requirements – Transcribed Data

- All of the Direct Input requirements
- Must be associated with the source data (e.g. paper)
- Must have undergone 100% data entry QC
- QA may want to perform a Risked Based auditing approach after the 100% QC.



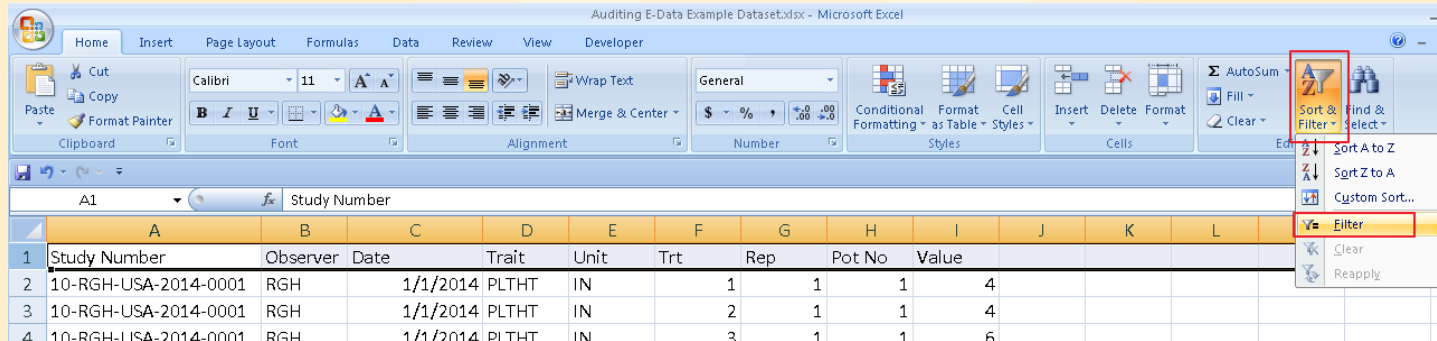
Excel Auditing Basics

- The Excel file must be:
 - Current
 - Complete - Includes the desired data
 - Closest to the end user/deliverable
 - No empty rows
- Utilize the *filter function* in Excel to allow:
 - Quick selection of items of interest
 - Counts of the number of records
 - Ranges of data values
 - Quick review of text based items

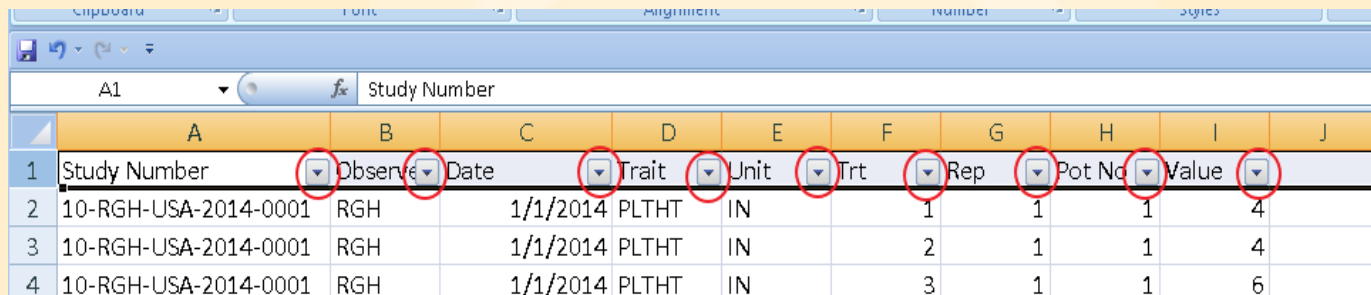


Applying the Filter Function

1. Select the top row containing the column headers
2. Select “Sort & Filter” from the menu ribbon and then select “Filter”



3. The filter is applied to each column



Applying the Filter Function

The screenshot shows a spreadsheet application window with a data table. The table has columns labeled A through I. The 'Trait' column (D) is highlighted, and a context menu is open over it. The menu includes options for sorting (Sort A to Z, Sort Z to A, Sort by Color) and filtering (Clear Filter From "Trait", Filter by Color, Text Filters). The 'Filter by Color' option is selected, and a sub-menu is open showing a list of traits with checkboxes: (Select All), PLTHT, SDWT, and STCNT. The 'OK' and 'Cancel' buttons are visible at the bottom of the sub-menu.

	A	B	C	D	E	F	G	H	I
1	Study Number	Observed	Date	Trait	Unit	Trt	Rep	Pot No	Value
2	10-RGH-USA-2014-0001	RGH			IN		1	1	4
3	10-RGH-USA-2014-0001	RGH			IN		2	1	4
4	10-RGH-USA-2014-0001	RGH			IN		3	1	6
5	10-RGH-USA-2014-0001	RGH			IN		4	1	7
6	10-RGH-USA-2014-0001	SEG			IN		1	1	15



Example Study

- Study: 10-RGH-USA-2014-0001
- Green house design:
 - 4 test groups
 - 2 replicates
 - 1 plant/pot
 - 2 plants/replicate
- Observer: RGH and SEG primary, FOH is occasional
- Observations:
 - Plant Height (PLTHT) in cm, 3 sets of obs, 2 plants/trt
 - Stand count (STCNT), 1 observation, 2 plants/plot
 - Seed weight (SDWT), 1 observation per trt/rep



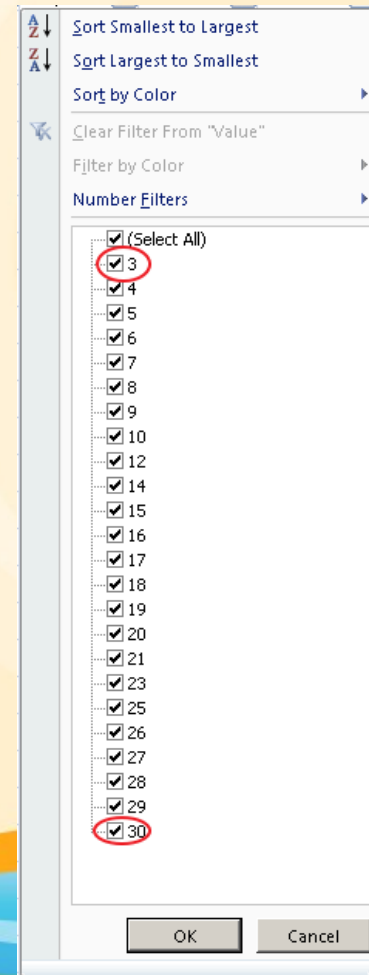
Example Cases

■ Plant Height (PLTHT):

- ❑ Base N: (4 test groups) x (2 reps) x (1 plant/pot) x (2 pots/rep) = 16
- ❑ Three collection dates: (3 Dates) x 16 = 48
- ❑ Units: cm
- ❑ Values make sense

74	10-RGH-USA-2014-0001	RGH	
75	10-RGH-USA-2014-0001	RGH	
76	10-RGH-USA-2014-0001	RGH	

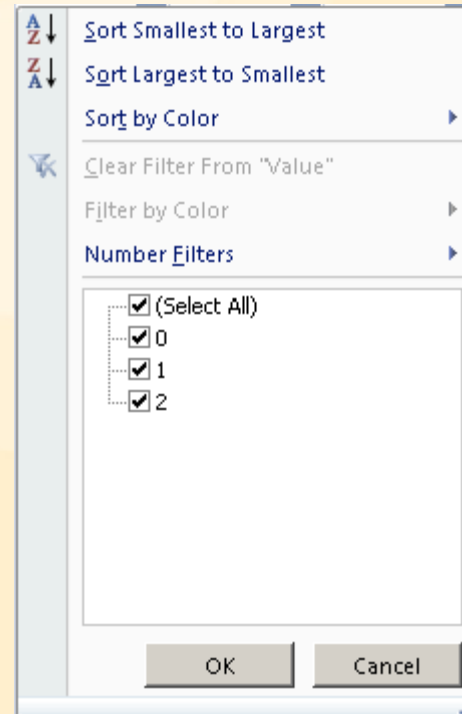
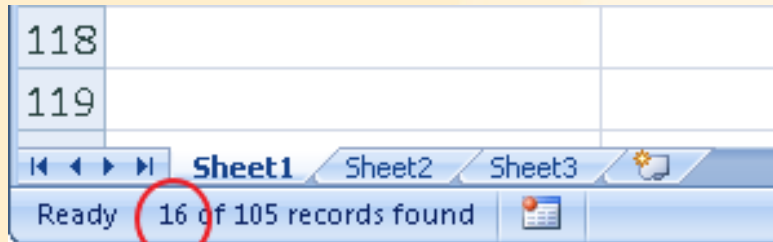
Ready 49 of 105 records found



Example Cases

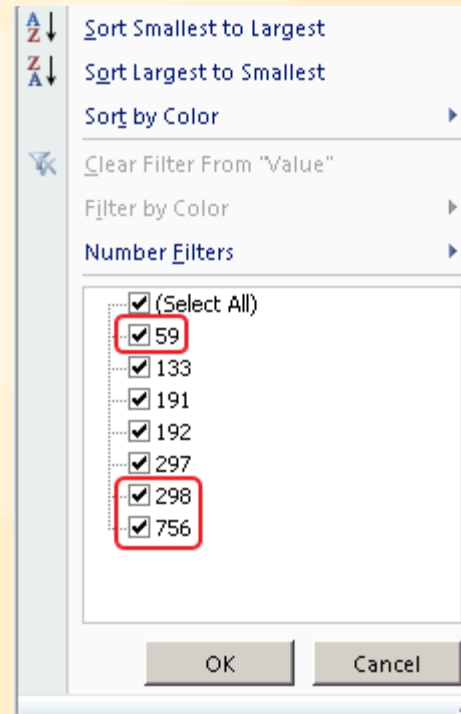
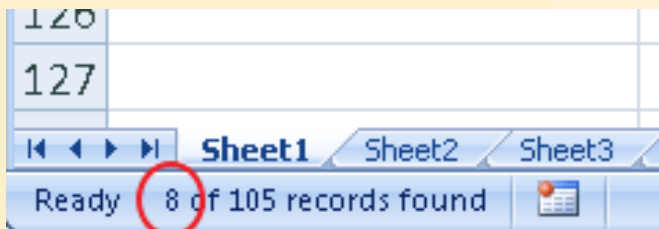
■ Stand Count (STCNT):

- Base N: $(4 \text{ test groups}) \times (2 \text{ reps}) \times (1 \text{ plant/pot}) \times (2 \text{ pots/rep}) = 16$
- Values make sense



Example Cases

- Seed Weight (SDWT):
 - Base N: (4 test groups) x (2 reps) = 8
 - Values make sense



Watchouts

- Is the file the appropriate one to be looking at (most current, etc.)
 - Save the e-copy along with audit report in case of questions in the future.
- How are extraneous comments handled?
 - ...those comments that would normally be made on the paper data in the margins
- When doing text based filters – be carefull with upper vs. lower case issues.
- What does a blank or “.” mean?
- If “stage” specific values are present, do they make sense?
 - e.g. growth stage, insect presence relative to growth stage
- Random subset selection



In Closing

- ❑ Excel is an extremely powerful tool for performing audits on electronic data
 - Allows the auditor to focus on study requirements
 - Thousands of records can be handled/manipulated with relative ease
 - Quickly identifies systematic errors
 - Finds “hidden” problems quickly

