

User Guide

The workbook provides a means for you to collect internal information from Excel workbook files. It can also compare two workbooks. It helps you to create an inventory of XLS files and to collect some statistics to assist in risk analysis.

Contents of this document

[Installation](#)

[How to check the Excel application settings](#)

[How to scan a drive or directory](#)

[Problems](#)

[How to automatically process password-protected workbooks](#)

[What the scan result columns mean](#)

[The Links worksheet](#)

[PivotTable and Database Query links](#)

[How to compare two workbooks](#)

[Notes and support](#)

[Differences from ScanXLS 2.3](#)

[Appendix: Excel Document Properties](#)

Installation

A ZIP file is supplied with the spreadsheet and license file. The license file is not plain text, it is required by, and must be in the same directory as, ScanXLS3. Launch ScanXLS3 and enable macros (In Excel 2007, 'Active content') for it to operate. The spreadsheet is supplied in the format of either Excel 97-2003 (.XLS) or Excel 2007 (.XLSM) as ordered.

In Excel 2007, click the Office button, Excel Options, Trust Center, in the Macro settings section, enable "Trust access to the VBA Project model".

In previous versions of Excel, choose Tools > Macro > Security, in the Security dialog select the Trusted Sources tab, and check the 'Trust access to Visual Basic Project' checkbox.

How to check the Excel application settings

Select the worksheet **Excel** and click the **Run check** button.

This will tell you if there are any Excel settings that should be changed for greater safety.

Every setting has text in column C that is hidden if the setting is as it should be, otherwise it is made visible. You can click on a cell in column C to read its contents in the formula bar even if it is hidden.

This also lists the installed Add-ins, any custom toolbars, and any mapped network drives.

How to scan a drive or directory

a) List the spreadsheet files

It is recommended that you open the original SCANXLS3 file read-only and then before proceeding Save As a different name in the SAME directory as your license file (SCANXLS3.LIC). When you open the file, it should display your license information.

Select the sheet **ScanXLS**.

Click the Browse button or type the name of the starting directory into cell A1.

Click the **List Files** button. All XL* files in that directory and its subdirectories, even hidden ones, are listed starting in row 5. At the end, a message captioned **File List Completed** tells you how many files were found, and how long it took.

You can cancel the file listing by pressing the Esc key.

You can repeat the **List Files** for any number of drives, or edit the list of files. The listing accumulates unless you opt to clear the list when asked.

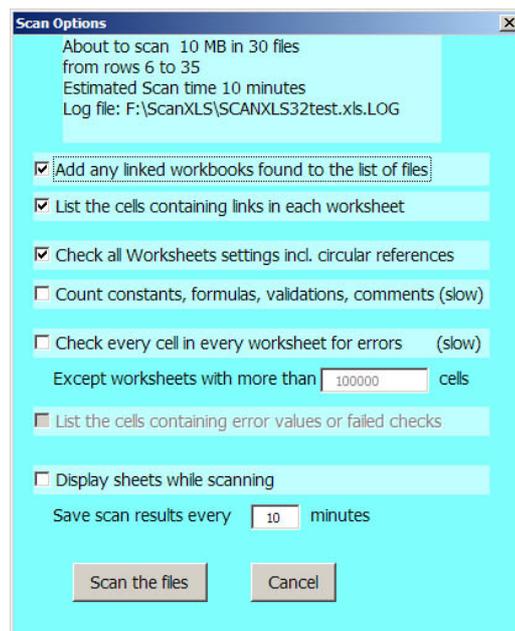
Review the list of files first and determine if there are any which you do not want analysed, for example any found in the Recycle Bin directories. To have the scan skip a file, hide or delete its row.

b) Scan the files

When you are ready to scan, first select the starting row - usually cell A5, and save the workbook.

Click the **Scan Contents** button. If you have not saved the file, you are prompted to save it under a file name generated from the current date and time, eg SCANXLS20070412160626.xlsm

The **Scan Options** dialog appears:



The log file is a plain text file that lists each file as it is scanned. If Excel crashes, examine this file to find the name of the last file it was working on.

The following are the effects of selecting those options:

Add any linked workbooks found to the list of files

If a link is found to a workbook that is not already listed in column A (Full Path) of the sheet SCANXLS, that workbook name will be added to the end of the list, so it will be included in the scan.

List the cells containing links in each worksheet

For each link found, the cells containing that link will be individually listed in sheet **Links** beginning in column I.

Check all worksheets settings incl. circular references

This reports any unusual options set for every worksheet in each file. Because checking for circular references requires doing a recalculation, this is optional so that if you know of any you can avoid external side effects during recalculation.

Count constants, formulas, validations, comments (slow)

It can take minutes to collect statistics on the count of formulas etc. in all worksheets in large workbooks so this may be skipped if you are really just interested in a fast scan, for example to check only for links.

Check every cell in every worksheet for errors

This counts both error values and cells that fail Excel's built-in error checking rules. This can take a long time. The results are displayed in sheet **ScanXLS**.

Except worksheets with more than [] cells

Worksheets with more than the given number of cells will not be checked in detail.

List the cells containing error values or failed checks

This is only available when the option above is selected. This option lists the addresses of the cells found in the column to the right of the count of errors. If this exceeds memory for collections of very large workbooks, do the scan with this option unselected.

Display sheets while scanning

This may be useful as a progress indicator. Leaving this unchecked will run a little faster. The status bar at the bottom left of the Excel window will still show what ScanXLS is doing.

Save scan results every [] minutes

This will autosave the workbook every so many minutes.

The scan works down from the starting row until it reaches a blank cell in column A. The file names in column B are changed to hyperlinks so you can easily open any given file by clicking on the link.

Each file is opened in turn read-only and displayed to help you observe progress:



Detailed cell checking can be slow so at times you may see no apparent activity. You can observe the scan progress in the Excel status bar. You may see, for example:

68% done of 59300 cells of sheet 'COAL', 9 of 13 in [temp.XLS], file 28 of 179 files , please wait

If you switch away to another task, and then switch back to Excel, the window may appear to be frozen with no status messages being displayed. Please be patient and wait for it to complete the current file, at which point it will refresh.

You can cancel the scanning by pressing the Esc key.

If ScanXLS is unable to close any workbooks after scanning them, you should close them yourself (without saving any changes) when the scanning is complete.

Problems

ScanXLS works in Excel, so any corrupt spreadsheet that it opens might cause Excel to stop operating or crash.

If you get an error message for an unrecoverable error such as the following, then just click OK.

- Device I/O error
- Unexpected error
- Microsoft Visual Basic System Error

If you get a 'Name Conflict', the workbook may have created in a double-byte (eg Asian) language but your PC has only single-byte (ANSI) support. Enter any name to resolve the conflict, and click OK. The file is opened readonly and will not be saved, so no harm will be done.

If Excel stops working or crashes, examine the end of .LOG file to find the name of the last file worked upon that may have caused the crash. When you re-open ScanXLS, the hyperlinked filenames tell you how far it got when it was last saved. You can then try the following:

- Reboot and then restart the scan with the selected cell in column A for the last file processed. If Excel had crashed after a memory leak, it might process it correctly the next time.
- Hide the row for the file causing the crash, thereby skipping the file during processing.

How to automatically process password-protected workbooks

If a workbook name matches a name in column A of the worksheet "Passwords" then the corresponding password in column B is applied to open it.

The program first tries to match the full path name of the file, and failing that it then attempts to match just the filename. That way, the same workbook with the same password can be in multiple directories.

Otherwise, the **Observations** column will simply show an error description.

If you don't want the Passwords sheet seen, hide it; or if you don't want to use the feature, delete the Passwords worksheet.

Please note that this only applies to passwords for opening workbooks, not for unprotecting worksheets. ScanXLS does not unprotect worksheets.

What the scan result columns mean and how to customize them

A red triangle in the top right corner indicates that a cell has a comment. Row 2 contains some summary statistics, either the SUM, COUNT, or MAX of the data below it.

Column	Meaning
Full Path	Full path and file name
Filename	File name only. Sort by this column to find possible duplicates.
Size	Filesize in bytes
Created	Date created
Accessed	Date last accessed
Modified	Date last modified
Attributes	A=Archive, H=Hidden, R=Read-only, S=System
File Format	The format of the workbook, usually the version of Excel it was saved as.
Observations	Any errors or other problems opening or closing the workbook or its referenced workbooks. You can enter your own observations into this column after the scan.
Scan Time	The time it took to scan this workbook.
User Notes	This is initially filled with the file sequence number. You may enter your own notes in this.
Property:	This heading is shaded green which indicates that you can customize it. Enter the name of a built-in document property after the colon. Examples supplied are: <i>Title, Author, Last author, Last Print Date</i> . A full list is given in the appendix at the end of this document.
Custom:	This heading is shaded green which indicates that you can customize it. Enter the name of a custom document property after the colon. The example supplied is: <i>Checked By</i> . A full list is given in the appendix at the end of this document.
Evaluate:	This heading is shaded green which indicates that you can customize it. After the colon, enter an expression that you want evaluated for the first worksheet. Examples supplied are A1 and B1.
Find:	This heading is shaded green which indicates that you can customize it. After the colon, enter the text that you want searched for in all worksheets. Examples supplied are IF(*IF(, INDIRECT(, OFFSET(, LOOKUP(, because these indicate a degree of model complexity. The column contains the count of cells found containing the text specified.
Where	Location of cells where the text in the column heading to the left was found.
Links In	Count of links FROM other workbooks found in this scan. This is particularly useful as there is no way to find that from the workbook itself.
Target Links	Full path and file name of workbooks using this as a source of data.
Links Out	Count of links TO other workbooks. The Links worksheet contains the same data, one file per row.
Source Links	Full path and file name of workbooks that this takes data from.
Calc Version	Version of Excel it was last calculated in; if zero, it needs to be recalculated.
nrUWB	Count of workbooks with unusual settings, see next for the list of settings reported. To learn more about what these settings mean, type the setting into Excel Help.
Unusual WB settings	"Accepts labels in formulas", "Create Backup when saved", "Custom Document Properties ", "Excel 4 Macro sheets", "Has Routing Slip ", "Is running as an AddIn" , "Multi User Editing", "Precision As Displayed " , "Remove Personal Information is enabled", "VBA code has been digitally Signed"

nrUWS	Count of worksheets with unusual settings, see next for the list of settings reported.
Unusual WS settings	Circular Reference, Consolidation sources, Filtered with hidden rows, OLE Objects, Pivot Tables, Protected, Query tables , Scenarios, Lotus evaluation rules, Lotus formula entry , Hidden, Very Hidden, Trailing space in name, Hyperlinks
Names	Count of range names in the workbook
Worksheets	Count of worksheets
Modules	Count of Visual Basic for Applications (VBA) standard modules
Forms	Count of VBA forms
VBA Code lines	Total count of lines of VBA Code in all modules
Validation	Count of cells with validation criteria
Comments	Count of cells with comments
Constants	Count of cells containing only constant (literal) values, excluding protected worksheets. This is not a check for constants embedded into formulas, see below.
Constant Errors	Count of cells containing only constant (literal) error values, excluding protected worksheets.
Numbers	Count of numeric input cells, excluding protected worksheets.
Formulas	Count of formulas, excluding protected worksheets.
Formula Errors	Count of formulas evaluating to # error values, excluding protected worksheets.
Cells	Total count of cells in the used area of all sheets in the workbook.
MaxCol	The highest column in use in the workbook, with its sheet name.
MaxRow	The highest row in use in the workbook, with its sheet name.
	The following error-checking tests are only available in Excel 2002 and later. These cell-by-cell tests can be slow with large workbooks.
NonEmpty	Count of cells that are not empty.
Overflow	Cells formatted as a date or time but with numeric values outside Excel date limits.
#Error	Count of cells with the error value shown in the headings #DIV/0! #N/A #NAME? #NULL! #NUM! #REF! #VALUE!
All Errors	Another count of errors; it should agree with the total of the #Error columns. The next column 'Where' then shows a list of sheets and cells with the error values.
Data Invalid	Count of cells failing their data validation conditions
Where	The list of cells found above.
Data Hidden	Count of cells where the data is hidden by font colour or custom formats like ;;;
Where	The list of cells found above.
Num in formula	Count of cells with a hard coded numeric digit in the formula.
Where	The list of cells found above.
Error Value	Count of cells evaluating to error. This may sometimes differ from that found in "Errors" above where there are multiple errors found by Excel's error checking. For protected worksheets this includes the count of cells where ISERROR() is true.
Where	This shows the addresses of the cells with those error values

Text Date	Count of cells with text contents that resemble a date
Text Number	Count of cells with text contents that resemble a number
Inconsistent	Count of formulas inconsistent with the region they are in
Omits Adjacent	Count of formulas that omit nearby cells.
Unlocked formula	Count of unlocked formulas
Ref to Empty	Count of formulas referencing empty cells
List Invalid	Count of cells in an Excel 11 Table (List) failing validation criteria
Where	The list of cells found above.
List inconsistent	Count of cells in an Excel 12 Table (List) with inconsistent formulas
Where	The list of cells found above.
Scoring	This is an arbitrary function that adds up the risk factors in the preceding columns.
	<p>Here is a reference copy of the formula in row 1 in case it is deleted:</p> $=LN(1+C1)+SUM(S1,U1,W1,Y1,AF1,AH1)+N(AA1)+N(AC1)+LN(1+AN1)+SQRT(AR1+AU1+SUM(BA1:BL1))+LN(1+BN1)+SUM(BP1,BR1,BT1,BV1,BX1,BZ1)+LN(1+CB1)+SUM(CD1,CF1)$ <p>Change it to suit your own purposes. Please note that no automatic tool can tell you the value at risk from the use of a spreadsheet. You need to examine how critical the decisions are which depend on this spreadsheet; you need to consider the controls, checks, and balances in the process around the spreadsheet; you should assess the skill levels of the developers and users using skills certification criteria such as ECDL.</p>

To keep the reports to a manageable size for a quick scan, all column entries are limited to 900 characters each. Any data that had to be truncated to fit is indicated by a chevron (») symbol at the end of long lines.

Multiline cell entries such as lists of cell addresses where errors were found, are also limited to 100 characters per line. These are indicated by three dots at the end of the line.

If you have workbooks with such numbers of errors, they should be examined with a detailed auditing tool such as EXChecker, SpACE, or Spreadsheet Detective.

External Pivot Table and Database / Web Query Links

Full Path	The path of the workbook containing the link
Filename	The filename of the workbook containing the link
Query Name	The name of the pivottable or query
Location	The name of the source data
Type	The type of link
Connection	The connection string for queries
CommandText	The query text

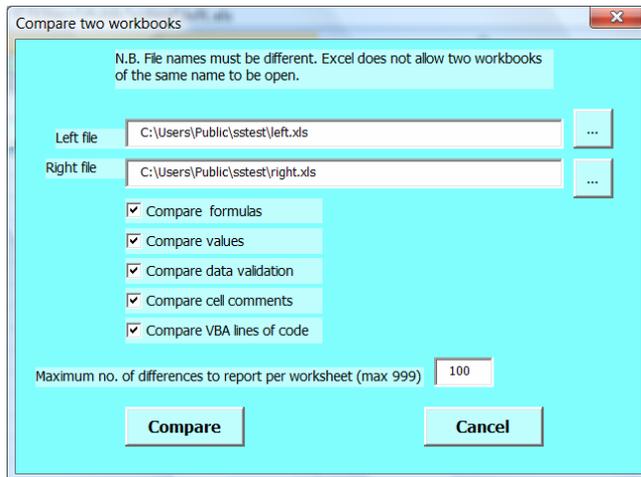
No.	Full Path	Filename	Query Name	Location	Type	Connection
1	F:\DOCS\SSample\wagovrace.xls	wagovrace.xls	resultsbycounty.aspx?o=300	'Governors Race'K1:N4	Web Query Table	URL:http://vote.wa.gov/general/resultsbycounty.aspx?o=3001&t
2	F:\DOCS\SSample\wagovrace.xls	wagovrace.xls	status_1	'Governors Race'J4	Web Query Table	URL:http://vote.wa.gov/general/status.aspx
3	F:\DOCS\ie2ke05	Using Data Validat	Query from MS Access Data	'Access Data'!A1:O73	Query Table	ODBC.DSN=MS Access Database.DBQ=c:\Work2000\ie2kexp
4	F:\DOCS\SDtest	Using Data Validat	Query from MS Access Data	'Access Data'!A1:O73	Query Table	ODBC.DSN=MS Access Database.DBQ=c:\Work2000\ie2kexp
5	F:\DOCS\SDtest	exampleleft.xls	FPNWIND	'Customer'!A1:K92	Query Table - OLE	OLEDB Provider=Microsoft Jet. OLEDB 4.0;Password="" ;User I
6	F:\DOCS\SDtest	exampleleft.xls	Flyingfitness	'Data'!A1:H41	Text Import	TEXT:E:\Flyingfitness.txt
7	F:\DOCS\SDtest	exampleleft.xls	Query from MS Access Data	'Access Data'!A1:O73	Query Table	ODBC.DSN=MS Access Database.DBQ=c:\Work2000\ie2kexp
8	F:\DOCS\SDtest	exampleleft.xls	Qry01_ResultsData_1	'Qry01_Results'!C4:D9	Query Table - ADO	This external data range was created programmatically and can
9	F:\DOCS\SDtest	exampleleft.xls	CurrencyRates	'Query'!A2:I66	Web Query Table	URL:http://moneycentral.msn.com/investor/external/excel/rates
10	F:\DOCS\SDtest	QueryMaster.xls	Qry01_ResultsData_1	'Qry01_Results'!C4:D9	Query Table - ADO	This external data range was created programmatically and can
11	F:\DOCS\SDtest	QueryMaster.xls	GrossBkgsQryResults_from	'Qry02_Results'!C3:C4	Query Table - ADO	This external data range was created programmatically and can

How to compare two workbooks

First, ensure that all the worksheets and the VBA code in the two workbooks are unprotected. It is a good idea to save the unprotected workbooks as two temporary files with different names which you can delete after comparing.

Select the **WBCompare** worksheet.

Click the **Run Comparison** button and enter the left & right file names and the comparison options:



The files are opened read-only for safety.

It enumerates all the worksheets in the left file and checks that they exist in the right file.

If they exist but in a different sheet order, it will report that.

It shows any worksheets in the right that is not in the left; they cannot be compared.

It reports any differences in the used range, number of formulas, or OLEobjects.

It performs the same check for code modules, reporting if the number of lines of code differ.

If worksheets or VBprojects are protected, they cannot be compared - it will report zero formulas or lines of code.

Then it proceeds to check every cell in every worksheet in the left against the corresponding cell in the right. There is no limit on this, so it may take a long time for huge Excel 2007 workbooks.

If there are rows or columns inserted or deleted, it may get out of step and report many differences. The only way out of that is to change one worksheet yourself to make it closer in structure to the other, save it under a different name and then re-run the comparison against this adjusted workbook.

For each 100 differences it finds in a given worksheet, it will ask you if you want to continue with that worksheet. That will save seeing thousands of differences, if all you want to know is whether they are different.

When completed, some workbooks may be left open because ScanXLS was unable to close them programmatically. You should close them yourself, without saving any changes.

Sheet	Cell	Left file	Right file
#1		About	About
#1	UsedRange	Sheet#1 ('About') A1	Sheet#1 ('About') A1:B2
#24		Query	Query
VBAProject		29 Components	29 Components
#1.Object		ThisWorkbook	ThisWorkbook
#21.Object		Sheet2	Sheet2
#22.Form		frmAddARecord	frmAddARecord
#23.Form		frmExplainFormulaDirectory	frmExplainFormulaDirectory
#24.Module	1 Lines of code	#24 ('Module1') has 7 lines of code	#24 ('Module1') has 8 lines of code
First difference at	Line 1	Option Explicit	'comment added in header
#25.Module	1 Lines of code	#25 ('Module7') has 247 lines of code	#25 ('Module7') has 248 lines of code
First difference at	Line 67	Let MortPymt1 = (Slsp1 - DnPytm1) * (((Rate1 / 12) * ((1 + Rate1 / 12) ^ (Time1 * 12))) / ((Rate1 / 12 + 1) * (Time1 * 12) - 1))	Let MortPymt1 = (Slsp1 - DnPytm1) * (((Rate1 / 11) * ((1 + Rate1 / 12) ^
#26.Object		Sheet21	Sheet21
#27.Object		Sheet3	Sheet3
#28.Object		Sheet22	Sheet22
#29.Object		Sheet23	Sheet23
CircRef	C7	Value [36.363636363]	Value [36.363636363]
CircRef	C8	Value [63.636363637]	Value [63.636363637]
CircRef	C15	Value [6.3636363637]	Value [6.3636363637]
CircRef	C16	Value [36.3636363637]	Value [36.3636363637]
EqPlus	C13	Formula [=C11-C12]	Formula [=C12+C11]
CondForm	C3	Value [300]	Value [77]
CondForm	C6	Value [600]	Value [377]
CondForm	C21	D.V. Formula [Between '100' .. '200']	D.V. Formula [Between '50' .. '250']
Indir	D3	Value [h4]	Value [H6]
Indir	D6	Value [North]	Value [West]
Indir	D7	Value [South]	Value [0]
		16 Differences found	

Example of Workbook Comparison results

NOTES

Do not delete any of the range names!

ScanXLS has been tested with: Excel 2007 Windows Vista, Excel 2003 and Excel 2002 in Win XP, Excel 2000 in Win 2000, and Excel 97 in Win 98.

Disclaimer: this spreadsheet is supplied as-is with no warranty. If it fails to perform as described above, Systems Modelling Ltd. will refund what you paid us for it. To assist us in our quality control, you should return the failed spreadsheet with a description of the error and stating the version of Windows and Excel used.

For support, please email Spreadsheet@sysmod.com

Your comments are welcome.

Patrick O'Beirne

(+353)(0)53 942 2294

Copyright © 2004-7, Patrick O'Beirne, Systems Modelling Ltd, www.sysmod.com

Differences from earlier ScanXLS 3 versions

1. Faster file directory listing
2. Options added in Scan dialog to omit some features for speed
3. Reports external data sources for chart series and listbox controls
4. Reports mapped network drives and UNC filenames for links
5. New Columns: 'User Notes', 'Overflow' and 'All Errors/Where'
6. Scoring column excludes weight for file creation date
7. The method of counting constants and formulas is slower but more reliable when there are more than 8192 areas on a worksheet.

Appendix: Excel Document Properties

Built-in	Custom
Title	Checked By
Subject	Client
Author	Date completed
Keywords	Department
Comments	Destination
Template	Disposition
Last Author	Division
Revision Number	Document Number
Application Name	Editor
Last Print Date	Forward To
Creation Date	Group
Last Save Time	Language
Security	Mailstop
Category	Matter
Format	Office
Manager	Owner
Company	Project
Hyperlink Base	Publisher
Number of Characters (with spaces)	Purpose
	Received from
	Recorded by
	Recorded date
	Reference
	Source
	Status
	Telephone number
	Typist