

Validating Excel-based Spreadsheets

Robert Ladyman
File-Away Limited



Why wouldn't you validate?

We don't validate Excel Spreadsheets
because...



Excuses, excuses

We only perform one-off calculations

- Not true if it's not “untitled”
- A well-known Monitoring Authority
 - 100% QC check
 - Clarification: must be validated if used for pivotal data handling



Excuses, excuses

We don't perform any calculations, we just store data

- “We don't perform calculations, so we'll store it in something specialized for calculations”
- You need to state what you do, not what you don't do (you don't store your data in Paintbrush).
- Excel is an active agent (more later)



Why would you validate?

GxP and its relatives

- Data Protection Laws

...and who wants to keep doing 100% QC



What type of 'system' is a spreadsheet?

- Level 1 systems
 - Off-the-shelf “Engines”: O/S, Oracle, Word, Excel itself (you don't tend to validate them)
- Level 5 systems
 - Bespoke
 - Strictly validated
 - Spreadsheets are programs+data and level 5, NOT LEVEL 1
- Applies to all spreadsheets, not just Excel-based ones



How do you validate?

Use the classic, standard methods



Our old friends

- URS – User Requirements Specification
- Version Control
- Testing
- Documentation (including the validation report)



URS

- For the *desired* system: what you want not what you have
- Field-values (how many, + or -)
- Calculations (standards?)
- If you end up with Excel, make sure you know what version you might use



URS - Audit trails

If your URS states that you need Audit Trails, Excel spreadsheets are probably the wrong tool

- Sarbanes-Oxley has resulted in more plug-in systems (e.g. Wimmer, RSME)

...of course, audit trails aren't needed if you don't store data



Version control

- MD5 sums
 - External to the file, best and simplest tool
 - What?
 - Why?
 - There is no such thing as 'bit rot'...
 - How?



Testing

- Test data
 - Valid and boundary checks
- Predict the results (plan)
- Evaluate the test results (validation report)



Documentation

Outputs

- Test results
- Programming notes – how does it work internally (you'd do that for a program)
- Don't forget the user manual and SOP



Existing spreadsheets

- Call it “version 1”
- Write a URS
- Create and record an MD5 before any changes
- Test against the URS and any other claims



What *can* (and *can't*) be done?

- Excel Controls

Set passwords to control:-

- Access (to open the spreadsheet)
 - To modify (otherwise there is read-only access)
 - To protect / un-protect (to program it)
- Message displayed if changes attempted
 - **BUT** - the Microsoft Excel Help states (in "About worksheet and workbook protection") "*[Excel passwords are] not intended to be mechanisms for securing data or protecting [data]*"
 - 7 minutes



Ah...but...

- “You CANNOT lock it down”
- “You CAN lock it down”



What else *can* (and *can't*) you do?

- Excel Controls
 - Set cell types (integer, date, number)
 - Date format can be yyyy-mm-dd
 - Error message for wrong type when entered...
 - BUT this can be overridden by pasting and by the delete key
 - AND watch out for those 'active' conversions (remember 12-3, 12/4?)



Excel is an active agent

Test Value	Integer format	General format	Number format (2dp)
12-3	“error”	12-Mar 12/3/2008	39519.00
12/3	4	12-Mar 12/3/2008	4.00
12-Mar	“error”	12-Mar 12/3/2008	39519.00
12/3/07	“error”	12-Mar 12/3/2007	39153.00
22/22/22	“error”	22/22/22	“error”

The test value was NOT prefixed with '='. Shaded are the correct output – only 6 out of 15 are correct



Built-in Tools

Use the built-in Excel tools

- Tools – Formula Auditing
 - Precedents
 - Dependants
 - Trace Error



An example spreadsheet

- Designed for Excel 2003
- Calculates the sum and mean for each pair of values (for a sample of some sort)
- Banner for “Valid” and “Invalid”
- Handout has ‘all’ the parts (you might not print that)
- Not just for show



Example Spreadsheet URS (the highlights)

- Must indicate missing sample ID in a row
- Must indicate missing value 1 and value 2 data
- Date must be no later than current date
- Must not store data
- Must display study number



How can we match the URS?

- Have an area for meta-data
- Have an area for data entry
- Have an area for validity checks
- Protect all other areas so that they cannot be modified
- Use entry-field types



Don't take MY word for it...

- EuSpRIG - European Spreadsheet Risks Interest Group

"Research has repeatedly shown that an alarming proportion of corporate spreadsheet models are not tested to the extent necessary to support Directors' fiduciary, reporting and compliance obligations. Uncontrolled and untested spreadsheet models therefore pose significant business risks. These risks include:

- Lost revenue & profits*
- Mis-pricing and poor decision making due to prevalent but undetected errors*
- Fraud due to malicious tampering*
- Difficulties in demonstrating fiduciary and regulatory compliance"*



Don't take MY word for it...

Study	Spreadsheets	Number with errors	% with errors
Coopers & Lybrand 1997	23	21	91%
KPMG, 1997	22	20	91%
Lukasic, 1998	2	2	100%
Butler (HMRC), 2000	7	6	86%
Total	54	49	91%

Ray Panko, University of Hawaii. Lawrence and Lee examined 30 project financing spreadsheets: all 30 had errors. Error rate: 100%



Resources

- Ray Panko:<http://panko.shidler.hawaii.edu/>
- More research to frighten you:
<http://arxiv.org/find/all/1/all:+spreadsheet/0/1/0/all/0/1>
- ...and more figures:
<http://www.isaca.org/Template.cfm?Section=Home&CONTENTID=3!>
- The example spreadsheet:
<http://www.file-away.co.uk/spreadsheets.html>
- MD5 software (check with your administrator)
<http://www.pc-tools.net/win32/md5sums/>

