# SPREADSHEETS AND SOLVENCY II

**Dean Buckner**
**Financial Services Authority**

**July 2010**

# Agenda

- **Solvency II is the new framework for insurance regulation in the European Union. It is currently scheduled for late 2012.**
- **The focus of the regime is on capital assessment integrated with risk management and the option to use models**
- **Good management of data is integral to this.**
- **I discuss**
  - The use of spreadsheets in data management
  - The risks to good data management posed by spreadsheets
  - control considerations

# I will not be talking about modelling

- Model risk is a separate issue (and an important one)
- I will be talking about the *data management process*: collecting data from disparate sources and combining and ordering it in a way that the model can use.

# Data operation

- **Data operation: Any point in the system where you ‘do something’ with data:**
  - Interpretation, formatting, alteration
  - Joining, restructuring
  - Aggregation, extraction, derivation
  - Merging, translation

# Spreadsheets in data management

- **There has been little study of the use of spreadsheets here**
  - They are extensively used at the interface between systems (particularly in risk management)
  - Studies of spreadsheet error have generally focused on the use of spreadsheets as ‘whole system’
  - Not as ‘part of system’.

# Summary of the problem

- **The principle of data system design has been understood since the 1970's**
  - E.g. Codd 1970
- **The same principles apply independently of system implementation (Codd 1970)**
- **These principles are (often) correctly applied within individual data systems.**
- **They are nearly always incorrectly applied across groups of interconnected systems.**
  - With spreadsheets

# History of the problem

- Before the 1980's corporate systems were developed as single large firm-wide data-systems.
- During 1980's and early 1990's, there was a rapid and widespread growth of relational databases
- This led to fragmentation of data systems, and consequent need for massive amounts of data to be transferred between systems.
- At exactly the same time, spreadsheets emerged as the main tool for managing data operations.

# How spreadsheets are used

- **Four main types of operation**
- **Those that should occur at the input stage, but don't**
  - Interpretation, formatting, alteration
- **Those that should occur within the system, but don't**
  - Joining, restructuring
- **Those that should occur at the output stage, but don't**
  - Aggregation, extraction, derivation (also formatting)
- **Those that should never occur at all.**
  - Merging, translation

# Who uses them

- **All these operations are used to transfer data between data systems, usually on spreadsheets**
- **A class of user emerged in the mid 1990's (spreadsheet jockeys)**
  - Highly skilled in spreadsheet operations
  - Usually in middle-office and risk management
  - Good at spreadsheets, can be poor at thinking of overall data model.
  - Fanatically devoted to Excel, avoid more structured tools such as query language.

# Databases vs spreadsheets

- "… about 1988 or so, I discovered a database manager. It was like walking into shop class with an old hand saw and seeing someone using a motorized radial-arm saw for the first time. Now this was a tool. It was fast. It could handle huge amounts of material. It could make complex, precise cuts in seconds. You could make just about anything you wanted with one of these". http://www.bizjournals.com/phoenix/stories/1998/10/19/smallb4.html

- Power users have no problem working with Crystal Reports or Cognos because they understand database design, structures, relationships, joins, and SQL commands. To a casual user, on the other hand, this is all Greek to them. This is where the first point of frustration comes in: **casual users don't have the time to learn Crystal Reports or Cognos nor do they want to become programmers. (Aacounting Web, June 2010)**

## Key differences from ‘modelling’

- **Avoidance of mathematical and statistical functions.**
- **Concentration on**
  - String manipulation
  - Sorting (strings and values)
  - Filtering
  - Lookups
  - ‘Wide spreadsheets’

# Input-like operations

- **Interpretation**
  - When a user takes a freeform (unstructured) field, and interprets it as a structured one – this includes parsing of data fields (e.g. “USD.LIBOR.2Y”)
- **Formatting**
  - Reformats data in preferred way (e.g. 3.1.2010 to 1/3/2010)
- **Alteration**
  - Typically data cleaning

# Output-like operations

- **Aggregation (SQL 'Group By')**
  - 'sumif' is a favourite
- **Extraction (SQL 'Select … from')**
  - Tables are grouped by field using (Data / Sort), then saved and sent on
  - An SQL query from source system is used to download a CSV file to be loaded as a spreadsheet.
  - Opaque, hardwired or inaccessible queries are real danger
  - So is inappropriate use of Excel *indirect* function.
  - As is *autofilter*
- **Derivation**
  - Complex transformations are performed on spreadsheet to be uploaded later. E.g. matrix multiplication.

# Query-like operations

- **Joining**
  - My pet peeve.  The equivalent of a ‘join’ operation in SQL.
  - Usually done by ‘vlookup’
  - Fast efficient and safe if correctly designed using query language
  - Deadly and error-prone in spreadsheets
- **Restructuring**
  - The Excel ‘pivot’ function is usual suspect here

# The Big Join

- **Joining is essential to all risk management data operations**
- **They all have *The Big Join***
  - ‘Contractual data’ (derivatives transaction, annuity policies)
  - ‘Observational data’ (market price history, mortality tables)
- **This is always a complex operation, usually involving translation as well (the relevant key for the contractual data will nearly always be different from the key for the observational data).**

# Operations that should not happen at all

- **Translation (aka ‘mapping’)**
  - a term which has a meaning for one system is translated into a different term that has the same meaning for another system
  - Use of ‘mapping table’ or algorithm.
- **Merging**
  - Identical data sets from different source systems are joined together – ‘cut and paste’
- **These *should* not happen, but are *unavoidable* in a fragmented system.**

# Solution

- **I still see audit reports or project plans that recommend replacing spreadsheets and manual processes with 'IT solution'**
- **This will never happen**
  - It is impractical to replace 2 or more fragmented systems with a single system
  - Replacing the spreadsheet operations with 'IT designed' ones only compounds the problem and removes any ability of users to address problems.
- **The only 'solution' is to eliminate the worst processes, and to apply appropriate controls to the ones that remain.**

# Obligatory brief horror story interlude

- […] a reconciliation spreadsheet which compared the trade records within the main trade processing system with those in the general ledger system. The spreadsheet had to load reports from the two systems, parse them, and construct pivot tables to make the comparison. This was automated to some degree, with a 400-line (recorded!) VB macro. But once the macro had run, heavy manual editing was needed to make good the deficiencies in the original reports. To put a set of controls around the process as it existed would have made no sense: in effect, the only way to test the process would have been to produce a duplicate report independently.
- "Controlling End User Computing Applications" Jamie Chambers and John Hamill (Eusprig 2008) http://arxiv.org/ftp/arxiv/papers/0809/0809.3595.pdf

# Possible controls

- **User understanding of 'good data principles' is key.**
  - As is an understanding of 'worst practice'
- **Replacing users' 'spreadsheet concept' of data with a relational model would be helpful**
  - But unlikely. Perhaps better to develop data operation tools that combine best spreadsheet practice with an underlying relational model
- **Firms should carefully monitor use of spreadsheets in data operations**
  - Particularly with regard to the type of operation. How many key spreadsheets are using vlookup/sumif and so on?

## Other controls

- **Firms may consider group-wide standards (e.g. ISO) on data format.**
  - Resolving pointless formatting problems
  - E.g. decide between 02Y, 2Y, 24M etc.
- **Data stress testing?**
  - What happens to the final number if key data sets are inaccurate or incomplete?

# Solvency II progress

- **Data management thematic completed June 2010.**
  - Most firms recognise the need to develop data management framework.
  - Many recognise the need for thinking about the whole data process
  - Some have supported this by developing a comprehensive data policy.
- **FSA will be meeting firms over the next year as Solvency II implementation progresses**
- **FSA cannot dictate good practice, but can recommend things for 'firms to consider'.**

# Questions & Comments