# EuSpRIG 2009 conference report

The key to excellence is doing it right first time; and that was the recurring theme in the papers at the tenth annual conference of the European Spreadsheet Risks Interest Group held at ENS Cachan, Paris, on 2-3 July 2009.

**Jocelyn Paine www.j-paine.org, Documenting Spreadsheets with Pseudo-Code: Cash-Flow and Loans**
His theme was "First document, then Excel". He specifies time based models in plain text pseudocode and his application (Excelsior) generates Excel files. It reminds me of Angus Dunn's concept, see below, and commercial applications like ModelSheet.

**Day 1 Keynote: Martin Erwig, Oregon State University (OSU), Software Engineering for Spreadsheets: Challenges and Opportunities.**
He spoke of type checking, and described research that indicated that combining syntactic label checking with semantic dimension analysis finds even more faults. A real-world tool from RedRover software, an OSU spin-off, features goal-directed debugging. He mentioned Gencel and ClassSheets, tools for the generation of correct spreadsheets, which may be relevant to the work of Jocelyn Paine (above) and Angus Dunn (below).

**Day 2 Keynote: Deniz Sumengen, FRC**
She spoke of the Technical Standards for Modelling, from the Board of Actuarial Standards of the Financial Reporting Council. They welcome comments on the exposure draft available from http://www.frc.org.uk/bas/publications/pub1986.html

**Patrick O'Beirne Systems Modelling, Checks and Controls in Spreadsheets**
This talk went back to basics on structure: have cross-foot totals, SUMs that are safe for the insertion or deletion of data rows, self-checking formulas, and ended with the top ten questions on spreadsheet use.

**Susan Allen, HBOS, Excel Modelling, Transparency, Auditing and Business Use**
Susan reported on how HBOS went from a restricted modelling package to an Excel-based system because the users preferred it. It features a separation between input, processing, and output; a formula walker to help understand dependencies; and change logging.

**Tom Grossman, Ozgur Ozluk & Jan Gustavson University of San Francisco, The Lookup Technique to Replace Nested-If Formulas**
This described how to replace successive IF functions, all doing the same kind of test, with a VLOOKUP formula. It received the highest rating from the delegates.

**Leslie Bradley & Kevin McDaid Dundalk Institute of Technology, Error Estimation in Large Spreadsheets using Bayesian Statistics**
This also drew on Software Engineering experience for defect estimation techniques. Briefly, when

you have a very large spreadsheet to audit, you can estimate the number of defects there might be based on how many you found so far. That way you might be able to stop the check earlier either because it looks very good or because it's too bad to be worth continuing.

**Angus Dunn, Automated Spreadsheet Development**
Angus has a generator called RingtailXL for Excel project models. Components are specified using syntax similar to range names, documentation is included, versions and changes are tracked, and can be assembled into workbooks.

**John Hunt Excel for managers, An approach for the automated risk assessment of structural differences**
He described how DiffXL infers the nature of changes in block structure in order to estimate the riskiness of a change. For example, a simple move of a block would be low risk, formula changes could be high risk.

**Matthew Dinmore, University of Maryland Baltimore, Documenting Problem-Solving Knowledge: Proposed Annotation Design**
He described implicit, explicit, and literate notation. Implicit notation is inferred from the formulas; explicit is present in comments and textboxes; and literate programming focuses on the best way to describe the program rather than on its executable path. He presented an informed-explicit design consisting of a traditional multi-tabbed spreadsheet pane positioned next to a multi-tabbed document pane.

**Ruth McKeever, Kevin McDaid & Brian Bishop Dundalk Institute of Technology, Analysis of the Impact of Named Ranges on the Debugging Performance of Novice Users**
This won the Student Prize this year. The international panel of judges, including Martin Erwig of OSU, described it as a well-designed and thoroughly executed piece of research. This directly challenged the common advice to use range names. The results show that for novice debuggers "A spreadsheet that contains range names in formulas will be more difficult to inspect and correct than a spreadsheet that does not use names in formulas".

**Bill Bekenn & Ray Hooper Fairway Associates, Some Spreadsheet Poka-Yoke**
Poka-Yoke is a Japanese term meaning "Mistake avoidance". They gave advice with examples on why things fail when inserting, deleting, and copying structural elements, and how to design to avoid these common mistakes.

**Angela Collins BDO Stoy Hayward, Embedded spreadsheet modelling**
She is the BDO spreadsheet guru who provides on-the-job training; just in time; task oriented. It reminds me of Kath McGuire's talk from 2007, that training based on the tasks that people need to accomplish in their context is more effective than generic classroom-type training. She supported that by reporting on a survey of user satisfaction.

**Derek Flood, Kevin Mc Daid, Fergal McCaffery. Dundalk Institute of Technology, NLP-SIR: A Natural Language Approach for SS Information Retrieval**

This allows the user to type enquiries in ordinary English in order to query tables contained in a spreadsheet. Such techniques are available in database queries. It reminded Grenville Croll of the Lotus 123 addin called "HAL".

**Sriram Iyengar & John Shvirbely Medal.org and University of Texas, The Medical Algorithms Project**
He described this huge corpus of 11,000 spreadsheets that facilitate medical decision making. However there were no details of the quality check and control techniques that they exercised.

**Françoise Tort, François-Marie Blondel, Éric Bruillard ENS Cachan, From errors detection to behaviour observation: first results**
The paper described using Camtasia to capture actions by users examining a spreadsheet for defects, and later analysis. Two groups were used, Sciences and Social Sciences. Martin Erwig asked about the gender differences in the groups which might be an explanation of the different behaviours. They also presented their results of an experiment carried out at Eusprig with a digital pen, to learn how to capture timing of form filling.

**Étienne Vandeput Université de Liège, Belgium , Milestones for Teaching the Spreadsheet Program**
He described the common limitations in most spreadsheet training books and school courses – that they focus on the commands in the spreadsheet, not problem solving. This accords with the view of Angela Collins above. He is developing a corpus of problem-based learning examples.

**Grenville Croll Spreadsheet Engineering Ltd, Spreadsheets and the Financial Collapse**
He comprehensively outlined the risks: human error, fraud, overconfidence, interpretation, archiving, assumptions, opacity, reification, and enterprise inoperability. And the role of spreadsheets in the financial system: Ubiquity, Importance, Criticality, Legality, Centrality, Contagion. He pointed to statements from the UK Financial regulator indicating their knowledge of the risks in spreadsheets found in use. A possible lesson might be drawn from the collapse of the Jamaican banking system from over-reliance on uncontrolled personal spreadsheets. He was careful to avoid asserting causation - certainly all these CDOs could not have been developed and piggybacked into a dizzyingly unstable pyramid without spreadsheets. But the reason surely is greed; "don't look too closely as these things as long as they are making us money".

**Mel Glass EASA, Reducing the Risk of Spreadsheet Usage - A Case Study**
He gave a straightforward description of a method to web-enable a spreadsheet that gives full access to its features (eg VBA) which is more than what Excel Services does.