EuSpRIG 2017

Thursday 6th July 2017, Imperial Data Science Institute, Blackett Laboratory, Imperial College London, SW7 2AZ

08:30 – 09:00 **Arrival, Registration, Coffee and welcome**

09:00 – 10.30 **Session 1**

The Future of Spreadsheets in the Big Data Era, Imperial College, David Birch, David Lyford-Smith & Yike Guo

Alternative Spreadsheet Model Designs for an Operations Management Model Embedded in a Periodic Business Process, University of San Francisco, Thomas A. Grossman, Vijay Mehrotra, Mouwafac Sidaoui

10:30 – 11:00 **Session 2**

Structured Spreadsheet Modelling and Implementation with multiple dimensions part 1 modelling, Paul Mireault, SMMI International

11:00 – 11:30 **Coffee Break**

11:30 – 12:30 **Session 3**

Mitigating Spreadsheet Risk in Complex Multi-Dimensional Models in Excel
Steve Litt, Indigo Sun, Incorporated

Proposed Spreadsheet Transparency definition and measures, Craig Hatmaker

12:30 – 14:00 **Lunch and visits to the Data Observatory**

14:00 – 15.30 **Session 4**

The Reification of an Incorrect and Inappropriate Spreadsheet Model, Grenville Croll, EuSpRIG

Electronic Evidence and the Presumption of Reliability, Stephen Mason

Mitigating Spreadsheet Model Risk with Python Open Source Infrastructure
Oliver Beavers, Trivium Financail Group

15:30 – 16.00 **Tea break**

16:00 – 17:30 **Session 5**

Structuring Spreadsheets with the "Lish" Data Model, Alan Hall, Michel Wermelinger, Tony Hirst and Santi Phithakkitnukoon.

Edu-Edition Spreadsheet Competency Framework, Maria Csernoch, Piroska Biró, Univ. Debrecen

The Role of Spreadsheets in Clinical Decision Support: A Survey of the Medical Algorithms User Community, Simon Thorne, Cardiff Metropolitan University, The Medical Algorithms Company

17.30 **Conference closes. Food and drink for the evening will be found locally. Please come along and enjoy the post conference discussions. Details to be confirmed**

Session 1

The Future of Spreadsheets in the Big Data Era

David Birch[1*], David Lyford-Smith[2] & Yike Guo[1] [1]Data Science Institute, Imperial College London
[2]ICAEW IT Faculty

ABSTRACT

The humble spreadsheet is the most widely used data storage, manipulation and modelling tool. Its ubiquity over the past 30 years has seen its successful application in every area of life. Surprisingly the spreadsheet has remained fundamentally unchanged over the past three decades. As spreadsheet technology enters its 4th decade a number of drivers of change are beginning to impact upon the spreadsheet. The rise of Big Data, increased end-user computing and mobile computing will undoubtedly increasingly shape the evolution and use of spreadsheet technology.

*To explore the future of spreadsheet technology a workshop was convened with the aim of "**bringing together academia and industry to examine the future direction of spreadsheet technology and the consequences for users**". This paper records the views of the participants on the reasons for the success of the spreadsheet, the trends driving change and the likely directions of change for the spreadsheet. We then set out key directions for further research in the evolution and use of spreadsheets. Finally we look at the implications of these trends for the end users who after all are the reason for the remarkable success of the spreadsheet.*

Alternative Spreadsheet Model Designs for an Operations Management Model Embedded in a Periodic Business Process

Thomas A. Grossman, Vijay Mehrotra, Mouwafac Sidaoui, University of San Francisco

ABSTRACT

We present a widely-used operations management model used in supply and distribution planning, that is typically embedded in a periodic business process that necessitates model modification and reuse. We consider three alternative spreadsheet implementations, a data-driven design, a canonical (textbook) design, and a novel (table-driven) technical design. We evaluate each regarding suitability for accuracy, modification, analysis, and transfer. We consider the degree of training and technical sophistication required to utilize each design. The data-driven design provides insight into poor spreadsheet practices by naïve modelers. The technical design can be modified for new data and new structural elements without manual writing or editing of cell formulas, thus speeding modification and reducing risk of error. The technical design has potential for use with other classes of models. We identify opportunities for future research.

Session 2

Electronic evidence and the presumption of reliability

Stephen Mason, Chambers of Stephen Mason, Langford, Bedfordshire

ABSTRACT

In 1997, the Law Commission decided that there would be a presumption 'that mechanical instruments were in order at the material time' – a presumption that includes software code. This means that the prosecution can rely on evidence from software code being reliable and therefore does not require witnesses to authenticate the evidence adduced from software code. This has led to unfairness in legal proceedings. At the date of this paper, the presumption remains in place, and the legal profession continues to certify lawyers as being competent to practice without knowledge of electronic evidence.

Session 3

Mitigating Spreadsheet Risk in Complex Multi-Dimensional Models in Excel

Steve Litt, Indigo Sun, Incoporated

ABSTRACT

Over 750 million people use Excel [Infotron, 2017] (...that's literally 1 in 10 people on the planet!) Companies around the world leverage it for its power, flexibility and ease of use. However, spreadsheets are manually intensive and prone to error, making it difficult for companies to control spreadsheet risk. The following solution proposes to mitigate spreadsheet risk for a set of problems commonly addressed in a spreadsheet defined as "complex multi-dimensional models". "Complex" referring to certain types of applications that require functionalitysuch as sophisticated algorithms, challenging hierarchies and database write-back (i.e. planning, forecasting, etc.) and "multi-dimensional" referring to providing capabilities such as reporting, data input forms and ad hoc analysis on the different attributes associated with the resulting model. The solution is defined as a "PivotModel" because it works similarly to a PivotTable but is designed to leverage the robust capabilities of the Microsoft Excel platform.

Proposed Spreadsheet Transparency Definition and Measures

Craig Hatmaker

ABSTRACT

Auditors demand financial models be transparent yet no consensus exists on what that means precisely. Without a clear modeling transparency definition we cannot know when our models are 'transparent'. The financial modeling community debates which methods are more or less transparent as though transparency is a quantifiable entity yet no measures exist. Without a transparency measure modelers cannot objectively evaluate methods and know which improves model transparency.
This paper proposes a definition for spreadsheet modeling transparency that is specific enough to create measures and automation tools for auditors to determine if a model meets transparency requirements. The definition also provides modelers the ability to objectively compare spreadsheet modeling methods to select which best meets their goals.

Session 4

The Reification of an Incorrect and Inappropriate Spreadsheet Model

Grenville J. Croll, EuSpRIG – European Spreadsheet Risks Interest Group

ABSTRACT

Once information is loaded into a spreadsheet, it acquires properties that it may not deserve. These properties include believability, correctness, appropriateness, concreteness, integrity, tangibility, objectivity and authority. The information becomes reified. We describe a case study through which we were able to observe at close hand the reification of a demonstrably incorrect and inappropriate spreadsheet model within a small non profit organisation. The organisation continues to plan its future over the next 35 years using a spreadsheet based Discounted Cash Flow model with negative discount rates.

Structured Spreadsheet Modelling and Implementation with Multiple Dimensions - Part 1: Modelling

Paul Mireault SSMI international

ABSTRACT

Dimensions are an integral part of many models we use every day. Without thinking about it, we frequently use the time dimension: many financial and accounting spreadsheets have columns representing months or years. Representing a second dimension is often done by repeating blocs of formulas in a worksheet of creating multiple worksheets with the same structure.

Mitigating Spreadsheet Model Risk with Python Open Source Infrastructure

Oliver Beavers, Trivium Financial Group

ABSTRACT

Across an aggregation of EuSpRIG presentation papers, two maxims hold true: spreadsheets models are akin to software, yet spreadsheet developers are not software engineers. As such, the lack of traditional software engineering tools and protocols invites a higher rate of error in the end result. This paper lays groundwork for spreadsheet modelling professionals to develop reproducible audit tools using freely available, open source packages built with the Python programming language, enabling stakeholders to develop clearly defined model "oracles" with which to test and audit spreadsheet calculations against.

Session 5

Structuring Spreadsheets with the "Lish" Data Model

Alan Hall, Michel Wermelinger, Tony Hirst and Santi Phithakkitnukoon.
The Open University, UK and (4th author) Chiang Mai University, Thailand.

ABSTRACT

A spreadsheet is remarkably flexible in representing various forms of structured data, but the individual cells have no knowledge of the larger structures of which they may form a part. This can hamper comprehension and increase formula replication, increasing the risk of error on both scores. We explore a novel data model (called the "lish") that could form an alternative to the traditional grid in a spreadsheet-like environment. Its aim is to capture some of these higher structures while preserving the simplicity that makes a spreadsheet so attractive. It is based on cells organised into nested lists, in each of which the user may optionally employ a template to prototype repeating structures. These template elements can be likened to the marginal "cells" in the borders of a traditional worksheet, but are proper members of the sheet and may themselves contain internal structure. A small demonstration application shows the "lish" in operation.

Edu-Edition Spreadsheet Competency Framework

Maria Csernoch, Piroska Biró, University of Debrecen

ABSTRACT

Based on the Spreadsheet Competency Framework for finance professionals, in the present paper we introduce the Edu-Edition of the Spreadsheet Competency Framework (E^2SCF). We claim that building spreadsheet competences should start in education, as early as possible, and this process is a lot more effective if support arrives from expert teachers. The main feature of E^2SCF is high mathability computer-supported real world problem solving. This approach is based on – from the very beginning of training – a two-directional knowledge transfer, data and error analysis and handling, and the programming aspect of spreadsheets. Based on these features, E^2SCF is set up for basic and general users to build up firm spreadsheet knowledge and to develop transferable problem solving skills and competences.

The Role of Spreadsheets in Clinical Decision Support: A Survey of the Medical Algorithms User Community

Simon Thorne, Cardiff Metropolitan University, The Medical Algorithms Company

ABSTRACT

This paper presents and discusses the results of a small scoping survey of Clinical Decision Support System (CDSS) users from the Medical Algorithms Company website which hosts 24,000 different CDSS. These results are analysed, discussed, and compared with other similar studies and contribute to the wider understanding of how CDSS impact on clinical practice. The results show that CDSS provided by Medal are being used by clinical professionals in a variety of settings, both as an operational tool and as a research and reference tool. Whilst these tools are implemented and executed in a database, the initial logic is worked out on a spreadsheet. The paper describes that process and examines some of the results of the survey.