The delivery of regulatory reports has usually focused on the finished result. The process of compiling all the data for the final document has received little senior attention – as long as the reports are issued on time.

This perspective is being overturned by recent regulatory documentation. Whether it is the 2012 consultation on Risk Data Aggregation by the Basel Committee, the April 2011 report by the Fed/OCC on Model Risk Governance or the 2012 FSA Solvency II findings on the IMAP (Internal Model Approval Process), the approach to regulatory supervision is changing. It is no longer good enough to ‘just’ produce the final report. Executives must attest to the figures and be able to demonstrate that the processes to produce reports are robust.

Across the world the new swathes of regulatory reporting – such as Dodd-Frank, Form PF, Stress-testing, Solvency II and many more to come – are being accompanied by the regulatory questions of the kind ‘how do you know these numbers are correct?’.

So what is the problem? Surely IT systems enable reports to be quickly and reliably generated from central data sources? Unfortunately this vision is far from true.

Even when IT architectures show closely integrated systems, the reality is very different. Many of the data elements of a regulatory report will have been created via multiple manual operations. This manipulation and aggregation is typically spread across Excel files, CSV extracts and Access databases (increasingly called ‘end user computing’ or EUC) – almost all of which do not appear in the formal IT architecture maps.

With the new focus of regulators on the reporting process as well as the final report, all of these manual activities must be documented and demonstrably robust. This requires the ability to locate and audit the activity taking place in multiple EUC files. Indeed the control of EUC is specifically referenced in both the Basel and FSA reports mentioned above.

There is a problem, however. The control of individual spreadsheets only makes sense in the context of the overall data management process that they support. A business outsider (such as IT) may contemplate the implementation of spreadsheet control across file servers showing endless lists of Excel files. However, the business and data processes they co-ordinate are hidden in the minds of the

“It reminded me of a math lesson at High School where I was actually able to come up with the answer to a problem. When the teacher asked how I arrived at said answer, I pointed to my ink-stained back of hand, where I had done all the calculations. Unintelligible to all but myself, my lines of calculation failed to pass muster.... this vignette is how I’m starting to think about the way regulators are looking at risk and other reports from financial institutions. It’s not just about the answer anymore; the regulators want to understand the models underlying the reports and – more pertinent to us – whether the data used in the models is both accurate and consistent.”

Andrew Delaney, ReferenceDataReview.com
Introducing the term ‘data lineage’

In order to confirm the integrity of regulatory reports it is essential to know the journey of key data all the way from their sources, whether from corporate databases or expert-based assumptions. It is for these reasons that the term ‘data lineage’ has become such an important term during the implementation of Solvency II and is becoming equally important in banking supervision. The presence of EUC processes in many data lineage analyses raises three key questions:

- What purpose does each EUC application serve – in terms of both business requirement and functionality (e.g. calculation and/or data manipulation)?
- How does the data arrive in/depart from the spreadsheet (e.g. links, connections, cut-and-paste)? Unfortunately, links between multiple spreadsheets and connections to other data sources can create an extremely complex web that is difficult and time-consuming to resolve manually.
- During business-as-usual operations how do you know that the spreadsheet and the data connections have performed as required? For example spreadsheet data connections (links) are particularly vulnerable to the transference of stale data.

ClusterSeven software provides rapid mapping and analysis of data lineage in EUC-rich environments. Automatic monitoring then delivers consistent validation as part of business-as-usual processes, providing a robust replacement for expensive, error-prone manual checks.

The ClusterSeven solution provides several key capabilities to map and maintain the integrity of data lineage, and identify potential vulnerabilities, including:

- The generation of data lineage maps by ‘spidering’ from critical spreadsheets to document the underlying data precedent tree.
- The analysis of individual connection details within the EUC estate to:
  - Identify links to source files held in locations such as local drives and temporary folders. These files may not be retained or updated in the course of normal business, potentially corrupting data audit-ability and the reliability of future data operations
  - Identify precedent links that cannot be followed to valid source file destinations (commonly called

“Where a bank relies on manual processes and desktop applications (eg spreadsheets, databases) and has specific risk units that use these applications for software development, it should have effective mitigants in place (eg end-user computing policies and procedures) and other effective controls that are consistently applied across the bank’s processes.”

Principle for effective risk data aggregation and risk reporting
Basel Committee on Banking Supervision, June 2012

“Ensuring good quality data management is a fundamental requirement to support the continued success of Canopius. The real power of the software is its ability to embed appropriate data- and function-checks as part of our normal business practices. It is difficult to know how this could be done effectively without using this type of technology.”

Mark Allen, Head of Business Information, Canopius, Lloyds Managing Agent
broken links)
- Expose security risks such as unencrypted enterprise system passwords embedded in spreadsheet data connections
- Alert the risk of stale data such as in linked spreadsheets where source (child) files have been modified, but destination (parent) files have not been modified.

- The analysis of the individual files within the data connection architecture to expose:
  - Files containing a large number or proportion of data cells as these may represent significant locations for manual manipulation of data, such as cut-and-paste, joining or aggregation
  - The existence of constants (so-called magic numbers) in formulas that may represent undocumented assumptions or judgement
  - The presence of cell calculation errors that may indicate incomplete or erroneous data operations.

**Conclusion**

There is a new regulatory focus on the quality of risk and regulatory reporting. This is resulting in a demand to control those end user computing (EUC) applications that support the data aggregation for these reports.

To determine the appropriate nature of EUC controls firms should focus on the full data lineage that supports the end-to-end business process of data aggregation, rather than treating spreadsheets as isolated elements.

ClusterSeven enables the rapid analysis of data lineage in terms of vulnerabilities in the overall architecture and within individual links and files. ClusterSeven also delivers non-invasive monitoring of the data lineage architecture to embed efficient and effective control.

“Data quality and control have become critical to risk management in the current business environment. Spreadsheets lie at the heart of an insurance firm’s business and as such insurers need solutions to manage, control and authenticate their key spreadsheet files. ClusterSeven’s product and service capability are exactly what we needed to keep MAP at the forefront of best-practice operational risk management.”

Adrian Duggleby, Head of IT at MAP Lloyds of London underwriter