

## “ORIGIN OF SPREAD SHEET RISK AND IT STRATEGIES FOR ITS CONTROL IN ENTERPRISE MANAGEMENT”

DR HARIDAS S. ACHARYA  
PROFESSOR,  
AIMS, PUNE

### INTRODUCTION

The popular financial risk management news and analysis site, Risk.Net [ 3 ], attributes a significant proportion of Financial risks to use of spread sheets by the employees of the finance departments. Protivity Inc., a prominent risk and business consulting agency company confirms this and has presented an exhaustive analysis of the problem in its publication titled “**Spread sheet Risk Management, Frequently Asked Questions**” [1 ]. However most of the analysts seem to have missed one basic aspect of the problem namely the 'real cause of **Spread Sheet Risk**'. The risk can be attributed to the presence of **Unstructured Enterprise data** in almost all cases. Software solutions acquired by the companies are only capable of handling well structured data in most cases. The onus of preprocessing the data , whenever it is unstructured is on the employees of the organization, and require a lot of human intervention . Employees are left with no alternative but to fall back on tools like spread sheets in this respect.

The aim of the current article is to examine the real 'cause' of spread sheet risk from 'formal theoretical perspective' and to prove that the solution lies in 'managing the situation' from a integrated-data management approach and rather than treat the problem as a 'consequence' of 'indiscrete and careless use of spread sheets by untrained employees .

### SPREADSHEET AS AN E-DATA PROCESSING SUPPLEMENT : A HISTORIC PERSPECTIVE

Spreadsheet was and is a large electronic sheet that organizes data about transactions for a business person to examine. It spreads or shows all of the costs, income, taxes, and other related data on a single sheet of

paper for a manager to examine when making a decision. It summarizes information from many paper sources, reports from other more sophisticated business software, in one place and presents the information in a format to help a decision maker see the financial "big picture" for the company. **Spreadsheets** have been used by accountants for hundreds of years, the latest being their electronic forms. There are reasons to believe that Dan Bricklin of Harvard Business School as the "father" of the electronic spreadsheet. In 1978, Bricklin and Bob Frankston co-invented or co-created the software program VisiCalc. In fact Prof Richard Mattessich was the first to develop a computerized spread sheet in 1961. In 1980 with invention of DIF format, portability between different programs became a reality. Lotus 123 with its capability to prepare charts came up in 1983 and since then Excel, Appleworks, and Corel Quattro Pro have appeared and are being extensively used.

By the late 1980s many companies had introduced spreadsheet products. Spreadsheet products and the spreadsheet software industry were maturing. Microsoft and Bill Gates joined the fray with the innovative Excel spreadsheet. The spreadsheet entrepreneurs were moving on, so were the business individuals and Enterprises using spreadsheets were increasing at an alarming speed. Open source software world offered Gnumeric as a free, cross-platform spreadsheet program that is part of the GNOME Free Software Desktop Project. OpenOffice.org Calc and the very closely related LibreOffice Calc are free and open-source spreadsheets, also licensed under the GPL.

With the advent of advanced web technologies such as Ajax circa 2005, a new generation of online spreadsheets has emerged. Equipped with a rich Internet application user experience, the best web based online spreadsheets have many of the features seen in desktop spreadsheet applications. Some of them such as Office Web Apps or Google Spreadsheets also have strong multi-user collaboration features and offer real time updates from remote sources such as stock prices and currency exchange rates. Spreadsheets that can handle big data, are also a reality [5].

## WHY SPREADSHEETS ARE INDISPENSABLE IN THE ENTERPRISE WORLD

In 2013 Board walk tech had couple of key steps forward with new integration framework using its 'Super Merge' technology and advancements to configuring templates that are used for access and input. Both of which help further embrace and extend use of spreadsheets. In some cases, third-party or internally developed dedicated applications are the right choice. In others, embracing and extending existing spreadsheet-driven processes is the most practical approach. If your organization is currently using desktop spreadsheets for some collaborative business process, it probably is putting up with a host of issues that are the inevitable result of the spreadsheet's inherent shortcomings.

What can replace a spread sheet? Answer is 'Dedicated applications', either internally developed or third party solutions. Numbers in Table 1 throw light on interesting aspects.

Table 1 Results of Enterprise Benchmark Research (Source VENTANA RESEARCH)

Sr No	Reason	Measure (%)
1	Employee resistance to change	56
2	Replacement is too expensive	44
	Total	100
3	Business case not strong enough	50
4	Alternative could too difficult to use	31
5	Training requirement (to improve usage)	19
	Total	100

From the Enterprises perspective the key division happens to be

**Employee Resistance (56%) + Replacement is too expensive(44%) = 100.**

From the spread sheet users perspective the key division is

**Training Required (19%) + Alternative could be difficult to use (31%) + Business case not strong enough (50%) = 100.**

Why would any one say that the business case is not strong enough. The users must have some reason to say so, the reason becomes evident when we look at the constitution of data itself.

## ENTERPRISE DATA : ITS CONSTITUTION AND PROCESSING NEEDS

Most of the analysts who have attempted solutions to spread sheet risks have ignored the fact that indispensability of spread sheets has much to do with the very constitution of the 'Total Enterprise Data'. Fig 1 gives a graphical view of the continuum of the Enterprise data.

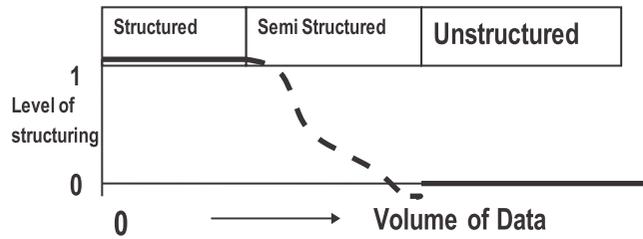


Fig 1. A graphic view of the CONTINUUM of Enterprise Data

The whole of an enterprise data would always constitute of three mutually disjoint data sets.

$$W = U \cup S \cup C$$

Where  $W$  is the Universe of discourse,  $U$  is the unstructured component,  $S$  is the Semi-structured component and  $C$  is the core structured component of the data. We may assume that these are mutually disjoint sets, or DOMs. Hence

$$U \cap S = \phi, S \cap C = \phi, U \cap C = \phi$$

The semi-structured component which is not properly handled by the dedicated applications is what is subjected to spread sheet processing by employees out of necessity. Replacing this usage by providing dedicated software could really be difficult. This is what is reflected in the fear 'business case not strong enough', as expressed by users when replacements are thought of. Thus it is evident that the very nature of data pose some basic Software Engineering Problems which need to be addressed prior to attempting elimination of spread sheet usage.

If one views the problem of data processing needs from the software engineering point of view we may conclude

1. The Component ( $C$ ) should be handled by well designed dedicated applications, and the processing will lead to Risk Free functionality.

Hence should pose no real problems.

*Without loss of generality we may assume that Enterprises have either procured or would be procuring well tested applications for processing of this component of data.*

2. Enterprise data would have a semi structured components. This component poses problems for proper Dedicated Application development as per the business rules.

*The semi structured nature of data makes the choice of analytic methods difficult and hence use of spread sheet becomes unavoidable.*

Attempt to reduce this component by using methods as discussed in [6] [7] [8] could help. Reduction of size of  $S$  and increase of size of  $C$  obviously will bring more and more functionalities under the dedicated applications which would be risk free.

3. The Unstructured component requires strong technology intervention.

*Currently the Enterprises have to hope for the technology to develop and solve the processing problems, and as such management by itself can do nothing.*

There is ample evidence to show that efforts are on to improve the situation. Many adhoc solutions seem to be existing and better solutions are forth coming ([2], [6], [7])

## TRENDS OF GROWTH OF SPREADSHEET RISK

The compound annual growth rates(CAGR) of data as projected by Industrial Development Corporation (IDC) estimates 21.8% CAGR in respect of transactional data, whereas the estimated CAGR for unstructured data is at 61.7%. There are no reasons to suspect the growth rates but not all the unstructured data is relevant to Enterprises. A better picture would be given by archived data, assuming that Enterprises would not archive data unless it is essential to them. Table 2 shows the distribution of total archived data from 2008-2015.

**Table 2: Relative sizes of archived U, C, S projected upto 2015 (Petabytes)**

Year	U	C	S	U%	C%	S%
2008	11430	1952	1652	76.0	13.0	11.0
2009	16737	2782	2552	75.8	12.6	11.6
2010	25127	4065	4025	75.6	12.2	12.1
2011	39237	6179	6575	75.5	11.9	12.6
2012	59600	9140	10411	75.3	11.5	13.2
2013	92536	13824	16796	75.1	11.2	13.6
2014	147885	21532	27817	75.0	10.9	14.1
2015	226716	32188	44091	74.8	10.6	14.6

Source of data : Enterprise Strategy Group, 2010, USA

A look at the trends shows that share of C is decreasing, share of U is also decreasing but share of S is increasing. Implications are very clear. A conscious effort to provide structures to unstructured data is possibly increasing the volume of semi-structured data. This would naturally put additional load on Spreadsheet type of processing in coming years, potentially leading to more Spread Sheet Risks. Invention of *Trillion-row spread sheet* [5] is a strong evidence of the fact that even big-data is going to be treated with spread sheet like solutions in the coming years.

### STRATEGIES FOR CONTROL OF SPREAD SHEET RISK

So far we have looked into the origins of spread sheet risk. Our focus has been on aspects which are related with the basic nature of enterprise data and software engineering aspects like to what extent lack of structure to data and resulting weakness in business case prevent dedicated application developments which can replace spread sheets. In conclusion it may be stated that control of Spread Sheet Risk is possible with three pronged approach.

Management must think *of reducing Spread Sheet Usage by increasing dedicated applications development*. This requires efforts to increase clarity in business cases, more efforts at providing structure to data so that software engineering becomes possible and affordable. This is an evolutionary process. Business rules keep changing, more and more unstructured data enter into business. The rate at which these things happen will always leave a gap between actual rate at which dedicated applications are developed and the rate at which demand increases.

Management must continuously *monitor the volume of Spread Sheet Usage, and properly audit the processes*. Incorporating additional checks at the very stage of designing the spread sheet solutions and strict quality control of the logic is necessary. In fact all the Enterprise spread sheets should be tested and validated as any dedicated application is tested normally. In addition, since designs of spread sheets can be changed by Employees at their end, the process of testing and validation should be a periodic event.

*Employees must be trained adequately*. They should be able to *distinguish between transaction processing and processing for analytics*. Multiple validations should be incorporated in case of spread sheets designed for transaction processing to reduce risk. In case of analytical usages *sensitivity analysis should be more stringent* so that managers who take business decisions based on the analytics are aware of the extent to which their decisions can go wrong.

Enterprises should accept the fact that spread sheet risk can only be managed and minimized there is only a very remote possibility that the risk can ever be fully eliminated.

### ACKNOWLEDGMENT

*Author acknowledges help rendered by Asmita Mer, Ritu Arora, and Vivek Kumar, students of Bharti Vidyapeeth, who helped specially in collecting information on Historic Aspects of Spread sheet Developments.*

### REFERENCES

1. Anonymous (2009) , *Spread Sheet Risk Management*, Pub Protiviti Inc., PRO-0609-101019, <http://www.protiviti.com/>
2. Anonymous(2013), *Global Enterprise Big Data Trends 2013 : Companes' current Tech. use and Big-Data plans*, Microsoft Inc., [www.microsoft.com](http://www.microsoft.com)
3. Anonymous(2014 ) Official Site for upto date financial risk management information for the global finance industry for risk management professionals, <http://www.risk.net/>
4. Anonymous(2010)*Trends in the Enterprise Storage Market*, <http://rvokal.fedorapeople.org/devconf2012/tomc-storage-trends.pdf>
5. Anonymous (2014) *1010data - Trillion-Row Spreadsheet - 1010 data - Big Data* , <http://www.1010data.com/solutions-and-services/self-service-analytics-for-big-data/trillion-row-spreadsheet/>
6. Dan Feldman\_ Melanie Schmidt† Christian Sohler (2014), *Turning*

*Big data into tiny data: Constant-size coresets for k-means, PCA and projective clustering*, Pub. Collective Intelligence, <http://www.scribd.com/collections/4379975/Big-Data>

7. Geoffrey Weglarz, (2004), *Two Worlds of Data - Unstructured and Structured*.
8. Peter Buneman, Susan B. Davidson, Many Fernandez and Dan Suciu (1997), *Adding Structure to Unstructured Data*, (in Lecture notes in comp. sci.), Int Conf on Database Theory, Vol 1186, pp 336-350, ([http://repository.upenn.edu/db\\_research/35](http://repository.upenn.edu/db_research/35))
9. Peter Hagggar (2011), *Data Growth and Standards*, Publisher IBM Inc.