Misuse of Spreadsheets in a Mission Critical Data Environment

To a man with a hammer, everything looks like a nail. **- Russian Proverb**

The universal temptation to grab the nearest tool at hand to quickly finish a household task can result in a long and ugly trail of splintered wood, stripped screws, mutilated hex heads bolts and bent, scarred metal. The responsible party can usually be located by following their howls of pain and outraged cursing. Look for the skinned knuckles and the pile of empty beer cans. The obvious lesson here is *there is no substitute for using the right tool for the job.* Having the knowledge and patience to go and *get* the right tool is the crucial factor. Surprisingly, this applies to computer software as well as hand tools!

A general purpose spreadsheet such as MS Excel is the "hammer" of software. Anyone can pick it up and start banging away on a problem and make rapid progress. Those friendly spreadsheet cells are always happy to accommodate some new data typed in, and the more the merrier! Add some equations and maybe a cool graph and you are really rolling! But rolling towards what? The unhappy fact is you are rolling towards big trouble for your enterprise.

Spreadsheets were designed as general purpose, ad hoc tools to analyze numbers, usually financial data. Spreadsheets *excel* at this task of ad hoc analysis. Columns of data, added or subtracted, moved around at ease, copied and pasted, etc. They are a fantastic tool to rapidly develop analysis scenarios and change them. It is just so easy. The trouble begins when this easy to use tool is adopted for critical tasks in the enterprise. Multiple users with varying degrees of skill and authority are left to interact with this open, easy to use, ad hoc tool. And chaos ensues.

The very features that make spreadsheets attractive to the casual user are the traits that make them very, very dangerous to the enterprise. Nothing is secure in this model. The data and equations just sit naked in the worksheet cells. Open. Vulnerable. Easy to change on purpose, and even worse, easy to change accidently. Drag and drop anyone? Whoops. Picture a real world scenario: *Damn! The phone rang just as I dropped the cell. It might have gone into the wrong column. I'm not sure. I didn't really notice where it went. Oh, well, there aren't any error messages and maybe my coworker will notice if anything is wrong. I'm going to email the spreadsheet to her. She needs to do some other editing and then email it back to me. I really hope there aren't any big mistakes. Hey look, it's time for lunch!*

Hope is *not* a very good security policy to protect your mission critical data.

Finding these errors is like finding a needle in a haystack. You won't find them. And the longer you use a spreadsheet, the more it is modified, the more people that will use it, the more complicated it becomes the more errors will creep in. It is insidious. Research studies have actually analyzed these error rates[1]. The statistics are far from reassuring. In fact, the statistics should horrify anyone using spreadsheets to store mission critical data, especially in an enterprise or multiuser environment. Bizarrely, the research also found users have a very high degree of confidence in the accuracy of their spreadsheets. The level of confidence actually *increased* as the spreadsheet became larger and more complicated! Hope springs eternal apparently. Needless to say, this overconfidence indicates a total lack of understanding of the threat.

Although much recent scrutiny of spreadsheet data integrity has concerned itself with the financial sector and the requirements of the Sarbanes-Oxley Act[2], exactly the same threats apply when dealing with any regulated environment at the municipal, state or federal level, especially any industry that has any impact on human health or safety. The storage, manipulation, and retrieval of critical data of any type, regardless of its nature, is an important job and the spreadsheet, although convenient, is a very poor tool to accomplish this job. Spreadsheets lack key features to protect data integrity. For instance, spreadsheets cannot track changes made to the data or the equations. There is no audit trail. When a

spreadsheet is passed onto another person for editing, there is *no* way you can tell what they did in their editing session. Old data may have been modified. New data may have been entered incorrectly. Equations may have been altered, perhaps one single operator in thousands of duplicated equations in row after row of data. Validation of data entry is clumsy and difficult to accomplish correctly.

Despite these vulnerabilities, spreadsheets are often seen as a low cost solution that justifies their shortcomings. This rationale may not be accurate however. The true monetary costs associated with the development and maintenance of spreadsheets are often hidden or obfuscated to the point where a true accounting is not available to the organization. For example, the work done to collate data and maintain the spreadsheet usually falls to the trained analyst (business or technical) who needed to analyze the data in the first place. This activity can become very expensive. One published estimate of median annual expenditures[3] for spreadsheet creation and maintenance was calculated at $780,000 per year just in analyst wages! The authors of that study indicate the mean figure would be significantly higher.

These and other problems inherent in spreadsheet based applications must be contrasted with applications designed using a database management system. Not only is data storage and retrieval of a much higher order of performance, it is much *safer*. Database management systems have built in features such as user accounts to restrict read and write access to the data. The manner in which the business logic is stored and structured prevents casual users from changing formulas/equations by accident or on purpose, and makes the logic much easier to develop, validate, and maintain. Support for user interface elements such as data entry forms provides a structured environment for getting data into the system safely and efficiently. Audit trails are common with database systems, allowing for tracking of every change made to the data by any user. For anything beyond casual use, the database is a much better choice for developing applications. For development of enterprise level multi-user applications, a database tool is the only prudent choice. Note that use of a database does not preclude the use of spreadsheets to *analyze* data. In this context the spreadsheet tool is used as a *window* to the data rather than the actual storage mechanism. This allows the spreadsheet tool to *excel* at its intended function while the backend database protects the integrity of the data store.

In this post 9-11 environment, the security and integrity of data in the enterprise is going to get increased attention. Keeping critical data in spreadsheets and *hoping* that there are no errors will be seen increasingly as *negligence*. Federal, state and local regulatory officials who have been educated about the threat spreadsheets pose to data security will take an increasingly dim view of their use in the enterprise. And rightfully so. Because using spreadsheets for mission critical data storage and processing is *asking* for trouble.

1. What We Know About Spreadsheet Errors. Panko, University of Hawaii
http://www.opssys.com/instantkb/article.aspx?id=10384

2. Managing Spreadsheets. P. Howard, Bloor Research
http://www.opssys.com/instantkb/article.aspx?id=10380

3. Strategies for Managing Spreadmarts. Eckerson and Sherman, TDWI