Dimensional
INSIGHT®

5 IMPORTANT CONSIDERATIONS IF YOU ARE USING SPREADSHEETS FOR BUSINESS ANALYTICS

WHY AND HOW TO TRANSITION TO BUSINESS INTELLIGENCE AS A BETTER ALTERNATIVE

Published: August 2013 | Revised: August 2017 | Version 3

CONTENTS

Spreadsheets are used extensively throughout organizations of all types, across multiple job roles and functional groups. They are easy to use for simple data analysis and simple calculations. They are a good tool for visualization of simple tabular data via charts and graphs.

Despite this widespread use and adoption, spreadsheets appropriately address only a small fraction of an organization's business intelligence (BI) needs.

In this paper, we will address 5 considerations that will help you understand the limitations of spreadsheets for true enterprise level reporting, analysis, and decision-support. We will review why adoption of business intelligence solutions like Dimensional Insight's Diver® Platform is changing the way leading organizations manage, report, analyze, and deliver information throughout the enterprise. And we will outline steps for transitioning your organization's spreadsheet use to a more robust reporting and analytic solution.

CONSIDERATION #1: **MANAGING RISK**

These days, an organization's reputation is every bit as valuable as its hard assets. Misrepresentations of corporate finances have landed several CEOs in the unemployment line or in jail and have driven their companies into financial dire straits. Whether a misrepresentation is intentional or the result of a spreadsheet error matters little in the eyes of regulators and prosecutors.

Spreadsheets are inherently error-prone. Even a small spreadsheet can contain thousands of links and formulas, which are generally coded as cell-based references. For example, a simple calculation such as =F11 + G232 + H2000 is hardly self-documenting, and being off by just one row or column returns significantly different results from what the creator of this spreadsheet formula intended. This alarming phenomenon has been well documented:

☑ Researchers at The University of Hawaii found that 88% of spreadsheets contained errors

☑ Coopers and Lybrand found 90% of all spreadsheets comprised of more than 150 rows contained errors

☑ KPMG found 91% of 22 spreadsheets taken from an industry sample contained errors[3]

Throughout the years, sensational spreadsheet fiascos have made headlines, underscoring that spreadsheet errors can result in significant adverse financial consequences.

*A spreadsheet error was implicated for facilitating a $6 billion derivatives trading loss at **JPMorgan Chase.** Investigators found that "the Value at Risk (VaR) model that underpinned the hedging strategy operated through a series of Excel spreadsheets, which had to be completed manually, by a process of copying and pasting data from one spreadsheet to another, and that it should be automated but never was."[2]*

*Prominent economists **Kenneth Rogoff and Carmen Reinhart** used spreadsheets for their much-publicized paper advocating fiscal austerity measures as the antidote to recession-plagued economies. Researchers trying to replicate their work discovered an*

important spreadsheet error that appears to call this assumption into question. Sadly, European policy makers had embraced Rogoff's and Reinhart's erroneous findings as economic dogma, arguably prolonging the European recession.[3]

__MF Global,__ a major global financial derivatives brokerage firm, declared bankruptcy in 2011 after several years of liquidity problems, trading losses, fines, lawsuits, and the collapse of its stock. About a year prior to its collapse, a consulting firm determined it needed to

improve the "end user computer tools such as Excel spreadsheets" that commodities brokers used to monitor risk and track the amount of money in its customers' accounts. Those upgrades were never made.[4]

__Barclays plc,__ a British financial services company, purchased significant assets of failed global financial firm Lehman Brothers after the latter firm collapsed in 2008. The specific assets Barclay was willing to acquire were detailed in

a spreadsheet that was sent to the bankruptcy court. That spreadsheet contained 179 hidden (instead of deleted) rows with toxic deals Barclays never intended to buy.[4]

A group of scientists in Australia found that one-fifth of more than 7,000 Microsoft Excel files in published genomics research had

errors dealing with gene lists due to a default setting that converts gene symbols to dates and numbers. [5]

CONSIDERATION # 2: SECURITY

Stored as files on individual computers, spreadsheets are frequently distributed via unsecure email, can be freely copied and modified, and may end up getting distributed to unauthorized individuals. Even password-protected spreadsheets can easily be compromised in seconds using inexpensive third-party tools.[6]

Spreadsheets have no built-in audit trail that establishes data provenance from raw data through to the final formula. Once a formula has been edited, no mechanism exists for recording who modified the formula and for what reason, and no archive exists of the original formula. Has the data in the

spreadsheet been appropriately updated with this month's results? Has it been manually "adjusted" by an individual? There really is no way to be certain. Few organizations are disciplined enough to mandate that users store spreadsheets in 3rd party version control repositories. While increased discipline can be applied to help mitigate security issues, it is reliant upon compliance of every individual user, and this also taxes worker productivity with additional, manual tasks.

In contrast, true BI tools offer auditability as a standard built-in BI tool. Since BI data models are typically located on secure servers instead of on individual workstations, version control

can be easily implemented and automated. An extensive security hierarchy lets administrators set access privileges from a data model level all the way down to individual data elements, preventing sensitive information from being accessed by unauthorized users. Since everyone in the organization is interacting

with a "single version of the truth," sensitive information isn't being transmitted via unsecure email. Authorized users have convenient browser-based access to relevant information presented in easy to digest tables, reports, and dashboards.

CONSIDERATION # 3: **END-USER SELF-SUFFICIENCY**

Although they are a ubiquitous tool, available to nearly all business users, spreadsheets are not particularly easy to use. Most business users within an organization have only limited spreadsheet modeling and use skills. The formula syntax for any calculation beyond simple algebra is non-intuitive. Advanced features like pivot tables, data connections, and macros are cumbersome. Most business users must lean on "power users" within their organization to perform any advanced analysis with spreadsheets.

Furthermore, spreadsheets are clumsy for updating data that changes regularly. The burden falls upon the user to manually recreate transformations each time the data is refreshed for consistent analysis of regularly changing data. Merging data from different sources is similarly difficult, time-consuming, and error-prone.

By contrast, for most users within an enterprise, BI tools are much easier to use than spreadsheets. Data extraction, transformation, and loading (ETL) are automated and scheduled to occur as needed, so the data is always

current. Automated ETL can occur at night when system resources are readily available, delivering updated data models and self-service interactive reports to users in time for the start of each business day.

BI tools are designed specifically for "ad hoc analysis", easily digging into the data from any angle. Users are free to simply point and click their way through the data to identify problems or opportunities for improvement. Click a summary figure and drill down to the detailed transactions that make up that summary. Change dimensions and get a 360-degree view of the problem.

Using ad hoc data discovery, a user in a hospital setting might discover the cause for a spike in hospital acquired infections or a sudden backlog of OR admissions. A supply chain analyst for a manufacturer might detect a downward trend in customer shipments or trends in equipment downtime, dig into the details, and determine the cause.

An overriding goal of BI is to disseminate information quickly and clearly to as many workers across your organization as possible to aid in business decision-making. Self-service dashboards and reporting and analysis tools facilitate and encourage data exploration by providing users with an environment that requires little if any training and quickly enhances their productivity. Drop-down lists for filtering data, single-click drilldowns, and menu-driven navigation provide users with powerful functionality that is intuitively easy to use.

Just as important, users are protected from the underlying formulas and scripts that create the data elements displayed on the dashboard. This prevents novice users from getting into trouble or inadvertently altering data elements.

Contrast this approach to information delivery based on spreadsheets. A much higher level of expertise and caution is required from end users. Staring at a spreadsheet, no obvious data exploration path stands out.

A central feature for enhancing productivity found in BI platforms like Diver Platform is the ability to click on a data element in a report, chart or dashboard and immediately see the details that make up that summary. We can think of this as seeing the "numbers behind the numbers." See Figure 1 to see an illustration of this.

Compare this to clicking on a number contained in a spreadsheet. What appears is simply a cryptic formula of cell references that show how that particular cell value was derived:
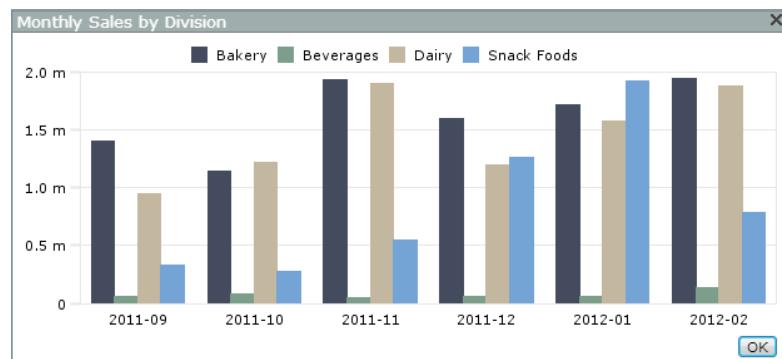
=SUM(C156:C105)-SUM(C58:C62

Figure 1. *BI provides users with powerful, single-click capabilities to drill down into "the numbers behind the numbers"—**not possible in a spreadsheet.***
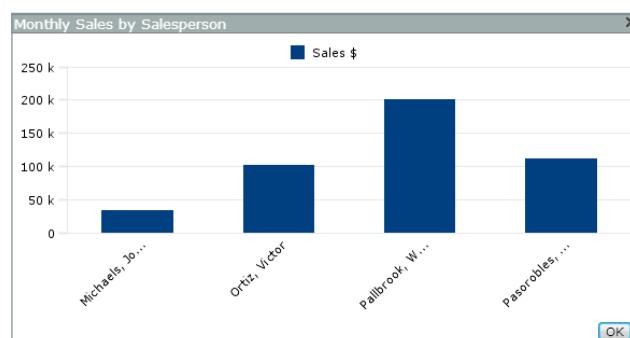
Starting from the top level Monthly Sales by Sales Group:
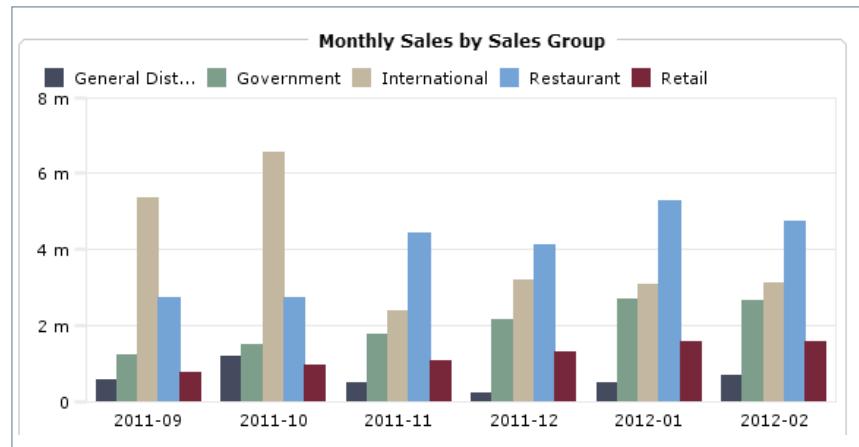
By clicking on a particular bar in the chart above, dashboard users can view Monthly Sales by Division:

...and then drill down even further, again by clicking on the bar of interest, to view Monthly Sales by Salesperson:

If spreadsheet users want to find the transactional details they will likely have to access a different spreadsheet tab or even one or more spreadsheets containing those values. Spreadsheet users are left to track down the source data and create their own data views, then create charts from that view. This is far more time consuming and error-prone than the single-click access provided by BI tools.

Imagine a scenario where everyone on your team, scattered across multiple geographic regions, wants to access the relevant sales metrics for their territory. Each team member is forced to create their own spreadsheet-based drilldown. This would be a serious misuse of sales resources, with less skilled users potentially creating the wrong data view. This scenario gets infinitely more complicated when data needs to be queried in real time or from several different source systems.

Little wonder that spreadsheet-based organizations spend exorbitant amounts of time arguing about whose spreadsheet is correct. Entire budgeting and forecasting meetings can be consumed by different parties arguing about whose version of the numbers is correct, leaving no time for making observations, collaborating to improve results, or agreeing upon corrective measures. BI platforms eliminate this unnecessary distraction. When team members operate from a single version of the truth, they can quickly and accurately analyze and understand the root causes of problems and drive change in the organization—a far more productive use of time than creating and debugging spreadsheets.

CONSIDERATION # 5: **BIG DATA**

Whether your organization deals with gigabytes or terabytes of data from a few source systems or hundreds, spreadsheets are at a distinct disadvantage for dealing with massive data volumes and disparate data sources and formats. Although spreadsheets can extract data from various sources and formats, glaringly missing is much of the necessary extract, transform, and load functionality found in BI systems (see Figure 2). This means that even simple SQL joins and merge operations have to be performed by a 3rd party ETL tool prior to importing the data into the spreadsheet.
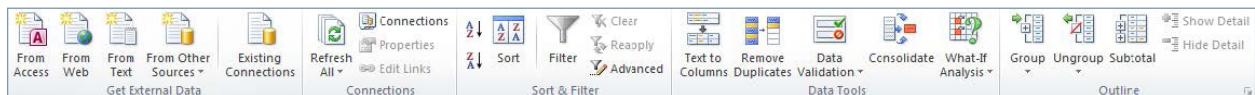

Figure 2: *The "Data" toolbar ribbon from Microsoft Excel 2010. Spreadsheets provide only the most rudimentary ETL functionality, unlike a full featured BI solution*

Contrast this rather limited set of functions with the rich set of ETL capabilities found in a BI tool (see Figure 3). Users can model complex ETL process flows from a point and click interface, and save this process in a repository for reuse. All defined ETL processes can be scheduled to run without manual intervention. Unlike the manual ETL processing with a spreadsheet, the BI platform generates detailed log files and audit trails. This capability is especially critical in ETL processes that handle sensitive HR, financial, or healthcare data.

Figure 3. Dimensional Insight's Visual Integrator 3.0. Robust ETL functionality lets users automate complex data handling workflows that would be difficult or impossible to perform using spreadsheets.
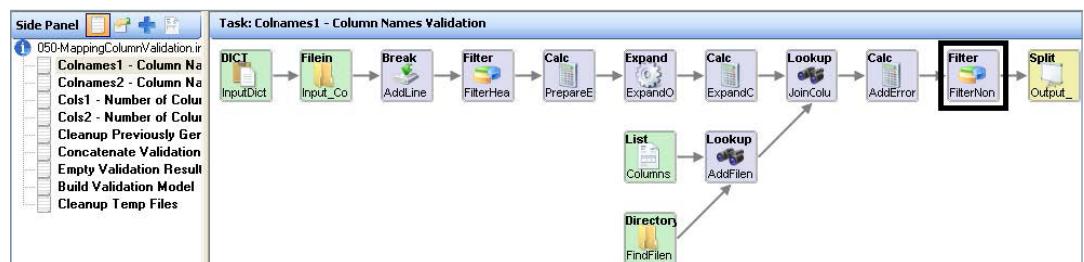
Spreadsheets were never intended to be used for multidimensional data analysis, and this becomes abundantly apparent as spreadsheet data volumes grow. BI tools return answers to complex queries quickly, with a few mouse-clicks (see Figure 4), because of the data model (or OLAP cube) underpinning a BI vendor's platform. These data structures are optimized for multidimensional data, which also explains why BI solutions can scale from megabyte to terabyte-sized data volumes. As spreadsheets grow in size, they become slow and unwieldy, resulting in lost productivity as users wait for queries to complete. Spreadsheet links and formulas are at increased risk of breaking, and ensuring formula validity becomes increasingly problematic if not impossible. Spreadsheets are designed to handle data across two dimensions. Analyzing more than two dimensions in a spreadsheet involves dealing with pivot tables or manually creating views of additional dimensions, which is a fragile, time consuming, and error-prone ordeal.

Reports that must be updated daily or more frequently also present significant problems when maintained as spreadsheets. Manual manipulation is required to append the latest data and delete or archive the oldest data. Then, new formulas have to be copied and validated to transform the raw data columns into the final data elements that are displayed in the report. BI tools generate detailed metadata, including timestamps that record the most recent updates. By comparison, spreadsheets have no built-in mechanism for notifying users whether or not their data elements and formulas are up-to-date.
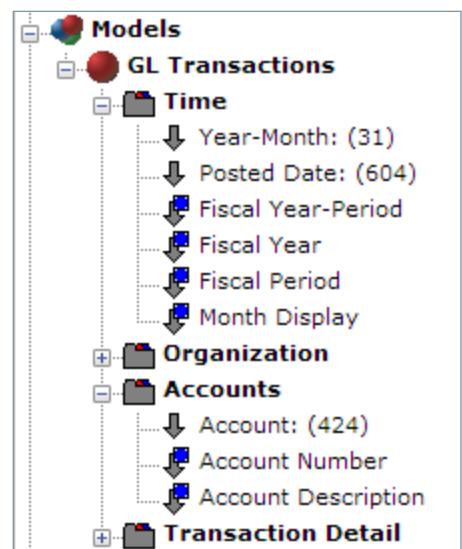

Figure 4. Dimensional Insight's NetDiver 6.4 data exploration console. Business Intelligence tools provide information workers with fast, single-click data exploration and analysis across any number of data dimensions.

NEXT STEPS: **TRANSITIONING SPREADSHEET USE TO BI IN YOUR ORGANIZATION**

Spreadsheets are not a business intelligence solution. Used for simple ad hoc calculations, spreadsheets are a useful and proven tool. Organizations using spreadsheets for purposes more appropriately addressed with a BI tool are exposed to reputational and financial risk, lost productivity, and significantly slower time-to-insight as information workers struggle with the inherent limitations of spreadsheet-based reporting and analysis. Here's how to approach the transition to true BI:

STEP 1. **Understand and embrace spreadsheet users in your organization.**

Strive to understand why they are using spreadsheets in order to uncover the types of data integration and reporting they desire:

❏ What groups are using spreadsheets regularly for advanced data integration, analysis, and business decision support?

❏ What data are they crunching? From what sources? What calculations are they performing?

❏ How are they charting and presenting the data?

❏ How frequently and regularly do they do it?

❏ Who are they circulating the spreadsheets to?

❏ What business decisions are being made based on the analysis?

❏ What shortcomings and pain points are they encountering?

STEP 2. **Immediately address any urgent concerns you discover.**

As noted earlier in this paper, the perils of spreadsheets are numerous and can have wide-ranging impact. If you uncover exposure to significant business risk or data security, take immediate action.

STEP 3. **Educate users about the perils and shortcomings of spreadsheets:**

Many business users are unaware of the security threats, liabilities, inefficiencies, and error-prone nature of spreadsheets. They are equally unaware of the advantages of BI as an alternative.

STEP 4. **Select a BI solution that works well with spreadsheets.**

Organizational acceptance of your BI solution will not be made easier by waging a holy war against spreadsheets. Spreadsheets are often an important component of an end-to-end BI solution:

☑ Be sure your BI solution can use spreadsheets as a data input source. For example, many organizations store important source data like annual quotas and bonus schedules for their sales staff in spreadsheets and nowhere else.

☑ Your BI reports and dashboards probably will need to have export capabilities to spreadsheets and/or PDF. It is simply the easiest way to enable users to print, share, and present data to others. This capability will be demanded by your users.

STEP 5. **Introduce BI where spreadsheets are most prevalent:**

Heavy spreadsheet use is a good indicator that BI can be adopted successfully in that group. It indicates that those people are hungry for analytics for business decision support, and they are most likely to embrace more efficient and reliable tools. Phased implementation of BI at the departmental level can be a cost-effective way to introduce BI into an organization, quickly establish success, and build a case for deeper penetration.

By migrating from spreadsheets to true BI, organizations can experience the following benefits:

☑ Industrial strength ETL functionality capable of reliably and consistently integrating significant data volumes from disparate sources

☑ Information delivery to end users accurately, quickly, and securely

☑ Powerful ad hoc analytics that don't require manual data manipulation or programming to get at the right numbers quickly

☑ Lessened liability and risk posed by error-prone, unsecured spreadsheets

☑ Increased opportunities to identify cost savings, efficiencies, and new business opportunities

References:

(1) "What We Know About Spreadsheet Errors," Raymond R. Panko, Revised May 2008, http://panko.shidler.hawaii.edu/My%20Publications/Whatknow.htm

(2) "Solutions to Spreadsheet Risk Post JPM's London Whale," Forbes, February 19th, 2013.

(3) "Reinhart, Rogoff, and the Excel Error That Changed History," BusinessWeek, April 18th, 2013.

(4) "Damn Excel! How the 'most important software application of all time' is ruining the world," CNN Money, April 17, 2013.

(5) "Gene name errors are widespread in the scientific literature," Genome Biology, August 23, 2016.

(6) "How to Remove, Crack, or Break an Excel XLS Password", http://www.online-tech-tips.com

ABOUT **DIMENSIONAL INSIGHT**

Dimensional Insight® is a leading provider of analytics, data management, and performance management solutions, offering a complete portfolio of capabilities ranging from data integration and modeling to sophisticated reporting, analytics, and dashboards. The company is a seven-time Best in KLAS winner in healthcare business intelligence and analytics, most recently ranking #1 in 2020. Founded in 1989, Dimensional Insight has thousands of customer organizations worldwide. Dimensional Insight consistently ranks as a top performing analytics organization by customers and industry analysts in its core market segments including healthcare, manufacturing, and beverage alcohol.

For more information, please visit www.dimins.com.

60 Mall Road
Burlington, MA 01803

t: 781.229.9111
www.dimins.com