



Spreadsheet and Database Risk Management

Lawrence Dunn, Steven J. Strauss, and Marcin Gwozdowski

May 2019

Contents

Objectives of the Whitepaper	2
Types of Spreadsheet Risks and Best Practices	2
Model Risk Management (MRM)	3
Procedure of Risk Mitigation	3
Alternatives to Spreadsheets – Databases	5
Types of Database Management Systems (DBMS)	5
Types of Database Storage	6
Business Intelligence – Definition	6
BI Pitfalls / Best Practices	7
Conclusion	8



Objectives of the Whitepaper

Presently, companies are heavily invested in spreadsheets because of their ability to aggregate, model, analyze, and report large data sets. Spreadsheets provide flexibility and familiarity, yet they are powerful and can handle a variety of tasks from the simple to the most complex. Spreadsheets can be an efficient and accurate tool when used properly. Setting up proper procedures and processes in advance is a crucial part of safely using spreadsheets while greatly reducing many of their inherent risks.

Types of Spreadsheet Risks and Best Practices

Across industries there has been increased spreadsheet use and reliance and, as a result, the integrity of spreadsheets has been under increased scrutiny. Headlines regarding spreadsheet errors, such as JP Morgan Chase losing over \$6 billion resulting from a copy and paste error and Barclays losing millions on contracts, they did not intend to buy resulting from hidden cells error, are more common than we realize and are willing to admit.

Spreadsheets can be prone to both human and design errors if proper processes and controls are not followed. Human errors, such as inputting or copying and pasting incorrect information, happen quite frequently. Designing spreadsheets to incorporate the proper logic in order to achieve a desired objective can also be complicated. Too much information on each tab, not labeling or mislabeling information, and using hidden cells can lead to unforeseen and undetected errors. Other errors that can also create problems include using dead or out of date links, inputs that are not properly marked or placed in the middle of a spreadsheet and therefore missed, and lapses in version control. One very useful tip to reduce this

error is to use feeder sheets rather than linking spreadsheets together.

There are several best practices which are commonly employed in order to prevent errors from occurring. To help limit human error, a company should ensure all users are competent and have undergone adequate training in the use of Excel, and enough time is designated for all users to review their spreadsheet(s). Periodic peer reviews should be established – often referred to as Model Validation controls. Allowing a different set of eyes to review and catch additional errors has proven to be very effective. Color coding input cells and formula cells can also lead to fewer input and copy/paste errors. Finally, including an instruction tab – How this Model Works - will ensure that anyone who opens the spreadsheet, whether they have prior knowledge or not, will understand how it was built, how it works, and how to modify it correctly, as the author intended.

To help mitigate the risk of design errors, the spreadsheet should not be over-complicated. By their nature, some financial and operational models may inevitably be complex. Breaking the operations of the model into building blocks will make the process of updating the spreadsheet simpler. Users should also limit the number of inputs and manual changes the spreadsheet requires for the same reason. Creating separate tabs designated specifically for inputs and assumptions will also streamline the spreadsheet's operation and function. Rows and columns should not be hidden, rather the "group" function should be utilized. This will dramatically limit copy and paste errors and make all cells easily visible. Confirming columns, rows, fields, etc. are properly labeled is of extreme importance. This clarifies the information used and ensures understanding of the output.

Model Risk Management (MRM)

In order to reduce spreadsheet errors a company can put a good Model Risk Management (MRM) system in place. By defining the purpose before building the spreadsheet, MRM ensures that all employees are aware of protocols and the need to document all changes and processes. Having protocols in place and documenting changes reduces the chance of material errors while increasing the understanding of the spreadsheet and its capabilities. MRM also provides user confidence in the spreadsheets' outputs and produces consistent and reliable results.

Procedure for Risk Mitigation

There are several steps in the process for risk mitigation.

- 1. The first step is to define a purpose for creating the spreadsheet.**
Defining a purpose ensures that each spreadsheet and its outcome is clearly identified. Part of this step is to involve multiple users in the ownership of each spreadsheet and determine who will design, maintain, and review the spreadsheet. These individuals should be competent and informed. They should provide periodic reviews or inspections of the spreadsheets. Another valuable addition to this step would be to add a peer review. A review by one who does not work on the spreadsheet regularly can lead to additional errors discovered. The final step for defining the purpose is to determine responsibility for timelines and record keeping. This person should establish a user guide that includes all controls and processes.
- 2. The second step in the procedure for risk mitigation is the design of the spreadsheet.** The flow of the spreadsheet should be clear and easy to understand by an outside viewer. The spreadsheet should move from left to right across columns and down the page. Using different tabs for different data sets is also imperative. Everything should be labeled, including the spreadsheet itself, each tab on the spreadsheet, each column or row used, and all data inputs and calculated outputs. Adding a separate tab, usually the first, for directions or the objective gives the user guidance. This tab should also contain all data sources, the data use, and all key calculations. To further enhance the design of each spreadsheet, multiple font types and sizes should be limited in use. Headings should stand out by being in a larger font than the remaining information as well as underlined or in bold font. Depending on the end-user's preference, printing should also be considered in the spreadsheet design.
- 3. The third step in the procedure for risk mitigation is managing the inputs in a spreadsheet.** The person designing the spreadsheet must consider how the data will be input. Separating the inputs from formulas will simplify the design. Adding a separate tab for just inputs and assumptions is highly recommended. Color coding formulas and input cells with different colors will also make each cell and its purpose easily identifiable. The last part of this step is checking the data. Using self-developed checks will help verify that each formula is working

accurately. Building checks that sum all financial information as well as sum counts that reconcile back to the original data can reduce unnecessary errors significantly.

4. **The fourth step in this procedure is managing formulas and functions.**

The user should make sure he/she understands how every formula should work. Users should also use filters that may be helpful in presenting the information needed. Filters should be used, properly applied and clearly identified. Tests should be set up for all formulas used. It is important to try to identify all outcomes for formulas and look for results that may show incorrect or incomplete values. These formulas usually result in a #N/A or #VALUE. Treatment for these errors should be researched in advance. Often an ISERROR formula will help deal with these situations and build logic into the formula to deal with these unwanted results.

5. **The fifth step is to maintain the spreadsheet.**

This step begins with clearly naming the spreadsheet and identifying the version or period covered. Good practice is to label or number the spreadsheets consistently using the same logic. Most of this step includes tracking changes and updates. Spreadsheets often evolve significantly during their life cycle and are modified by someone other than who originally built it. They are often designed for one purpose and are altered for other uses over time. This makes properly reviewing and documenting all changes critical. This step should usually be performed by individuals

independent of spreadsheet development.

6. **The sixth and final step in the procedure for risk mitigation is the location and security of the spreadsheets.**

The spreadsheets should be located on a secure network. Each spreadsheet should be saved in a defined part of the network. Everyone with proper access should know where each spreadsheet is and how to access it. The network should be set up automatically to back up the spreadsheets periodically as well as control or restrict unauthorized users from accessing the spreadsheets. The spreadsheets must have built-in security features as well. At the file level, the file can be encrypted which protects the spreadsheet from being opened unless a correct password is entered. The spreadsheet itself can also be encrypted. This allows the spreadsheet to be opened and viewed, but a password is required to modify any contents. Marking the file as final prohibits any further changes to be made to the spreadsheet. The structure and cell content can be password protected as well for additional security.

The process of risk mitigation needs proper financial and human resources to make it work correctly and efficiently. Examining the inputs, processes, and outputs will help identify which controls are most complex and most vulnerable to risk. This will help build a required level of control and procedures needed to create a framework to lower the risks (See Exhibit 1 on Page 9). Other items to consider which may enhance this procedure are establishing an Internal Audit Committee in addition to the peer review and using the most up to date

versions of the program. Risk mitigation is vital to an organization and with several simple and regular implementations, it can allow for fewer spreadsheet errors, leading to continued profit and success.

Alternatives to Spreadsheets - Databases

Occasionally, there is so much data that a spreadsheet is not the most efficient tool to use. In such circumstances, the correct tool is a database. A database is a systematic collection of data organized to allow easy retrieval when needed. It is a collection of data tables, schemas (the structure of a database), queries (a question about the data, expressed in a formal way), views (a searchable object defined by a query), reports (the formatted result of a query), and other objects. Databases are governed by Database Management Systems (DBMS), which are systems that handle the storage, retrieval, and updating of data in a database. A DBMS also helps to control access to the database.

Types of Database Management Systems

There are four major types of DBMS: Hierarchical DBMS, Network DBMS, Relational DBMS, and Object-Oriented DBMS.

1. **Hierarchical DBMS** employs a tree-like or parent-child relationship of storing data. Data is stored as records (a collection of fields with each field containing only one value) which are connected to one another through links. Hierarchical systems support databases where there is one parent and one or many children, known in data lexicon as “one-to-one” and “one-to-many” relationships. This model was the first type of DBMS

developed, with its roots going back to the early 1960s.

2. **Network DBMS** was designed as a progression from the restrictive Hierarchical structure. In the Network structure, object types are not limited to trees, and can support the “many-to-many” relationships that result in complex database structures. This model was developed in the late 1960s and is still used in certain applications today where there is high data volume, or where it would be cost prohibitive to move to a Relational DBMS.
3. **Relational DBMS** structures data storage through pre-defined relationships. Unlike Network DBMS, Relational systems do not support “many-to-many” relationships. Instead, they have several pre-defined data types that they can support. Relational systems are the most popular DBMS in the market. Examples include MySQL, Oracle, and Microsoft SQL Server.
4. **Object Oriented DBMS** stores data in the form of objects. These objects have attributes and methods that instruct the database what to do with the data through object-oriented programming (which tells the system what to do with the data). Object Oriented DBMS allows programmers to enjoy the consistency that comes with one programming environment because the database is integrated with the programming language and uses the same representation model. Certain Object Oriented DBMS are designed to work with

object-oriented programming languages like Java.

analytical processing (OLAP), alerts, and business analytics.

Types of Database Storage

Hierarchical and Network DBMS store data in “flat files” while Relational and Object-Oriented DBMS store data using defined relationships. Each type of data storage has advantages and drawbacks. Flat file storage puts all database information in a single table. It is simple to design, very portable and can typically be used without requiring robust software. Changing information in a flat-file storage structure is very easy as it has zero impact on any other file. The drawbacks of a flat file structure are that they are relatively slower, less robust than relational databases, and they are more difficult to use in sharing data.

Relational databases incorporate multiple tables to work together. The relationships between table data can be collated, merged and displayed in database forms. Relational databases offer more robust reporting through query tools that often incorporate Graphical User Interfaces (GUIs) that make designing queries very easy. The drawback of a relational system is that their structures can grow to be very complicated, which makes them difficult to set up and administer. Also, changing data in a relational system can be tricky. Unless there is a modeled means of updating records, making such changes in a vacuum can have disastrous implications on the entire relational model.

Business Intelligence - Definition

Business Intelligence (BI) is a rubric that includes applications, infrastructure, tools, and best practices that enable access to and analysis of information to improve and optimize decisions and performance. These components are broken out into five activities: querying, reporting, online

1. **Querying**, as mentioned above, is the activity of asking questions of a database in order to obtain a subset of the information stored therein.
2. **Reporting**, also referenced above, is the formatted output of a query.
3. **Online analytical processing (OLAP)** is a means for users to easily and selectively extract and query data in order to analyze it from different points of view (“what if” analyses, for example). The core of any OLAP system is a data cube, characterized by dimensions. The cube is typically composed of some type of schema in a relational database. Measured values are derived from the records in the “fact table” and dimensions structure are derived from “dimension tables.”
4. **Alerts** tools will send an email, text message or otherwise communicate with users once certain parameters or thresholds have been met in the monitored data set.
5. **Business analytics** focuses on statistics, prediction, and optimization of the reporting created by Business Intelligence systems/processes.

The output of Business Intelligence can be wide and varied. Some output are pedestrian spreadsheet files that contain raw data. Others can be formatted and summarized tables. Still others are robust graphical representations of the underlying data (often called Key Performance Indicators (KIPs) or Dashboards). To effectively use and set up BI tools, it is necessary to have the data drive the

reporting, and to have the reporting drive the data. Although this may seem counterintuitive, it is this bifurcated perspective that leads to the establishment of the most robust systems. While it is impossible to report on data that is not possessed, such that reporting must be modified to include only the data within the OLAP cube, that cube must also be modified to incorporate all of the data necessary to populate each reporting template created.

BI Pitfalls/Best Practices

Business Intelligence systems can distill huge amounts of complicated data into compact meaningful output. However, in order to do that they can be very complex. There are three significant pitfalls of BI that must be avoided to ensure consistent, high quality output.

1. **The first pitfall is not ensuring that there is a single version of the truth.** Several actions can lead to this. One such action is not putting proper controls on BI systems. If the BI system permits self-service, for example, multiple users can run the same report and get different results if they use different parameters. The best practices way to avoid this pitfall is to limit users' access. Let users run a report at their convenience, but only let them change one or two dimensions. Permitting users to change the time dimension (a date range) and an entity dimension (different departments) and locking down all other variables in a report is a great way to do this.
2. **The second pitfall is ignoring data storytelling to communicate and drive insight.** Anyone can run a report to calculate a metric. They can then share that report around

so that everyone is familiar with that metric. While the metric could contain valuable information, it is not the whole story. The best practices way to avoid this pitfall is by using BI tools, a skilled analyst can drill into the data and explain what is driving the metric, discuss any/all trends that are visible when looking at that metric over time, compare the metric with other benchmarks such as budgets/forecasts, prior periods, and finally, discuss what it means in terms of business strategy and impact. Failing to take this next step defeats the purpose of implementing BI tools to begin with.

3. **The third pitfall is not implementing as much automation as possible into the BI environment.** The entire purpose of BI is to streamline large data sets into small components of summarized meaningful information. Often, companies who impellent BI systems stop at the point where such reports can be created. But by taking the extra step to add automation, the value added can be increased exponentially. The best practices approach to automation is to use as much as possible, whether that is by automating distribution of the reporting, or by using an alerts tool to draw the audience's attention to certain metrics once thresholds are met/exceeded. Recently, automation has been taken a step further: machine learning and artificial intelligence have been incorporated into BI systems and processes to leverage computers doing what they do best – crunching huge volumes of data,

identifying patterns, learning from them, and generating recommendations to assist human users.

Conclusion

Spreadsheets are extraordinarily useful tools. They are dynamic, flexible, and can be used for multiple purposes. Because of these abilities, spreadsheets have become indispensable to many businesses, across industries, for data aggregation, modeling, analysis, and reporting. Like most tools, spreadsheets contain inherent risks; Without adequate oversight and appropriate controls, spreadsheets will produce inaccurate results, which can have disastrous consequences. It is for these reasons that spreadsheet risk that must be acknowledged, addressed, and mitigated through identifying a purpose, appropriate

design, managing inputs, using the correct formulas, spreadsheet maintenance, and applying appropriate security.

With large data sets, a spreadsheet may not be the appropriate tool. In such an instance, using a database as part of a database management system (DBMS) may be the correct solution. There are four types of database management systems, and they are used for disparate purposes. Regardless of which type is used, like with spreadsheets, there are pitfalls that need to be avoided to ensure quality output and data usage. But by ensuring there is a single version of the truth, making use of data storytelling and automating the DBMS, the efficiencies that can be gained are transformative.

Exhibit 1

Summary of “Best Practice” Procedure

