

# Prevent Database Nightmares

## EXCEL

### It's Not A Database

Spreadsheets can look very similar to a database table for non-database users.



## Excel - A Popular & Capable Spreadsheet Tool

Excel is a popular tool, and for a good reason. It's flexible, and it's easy to get started. This excellent spreadsheet tool can perform calculations and even create diagrams. If your organization uses Sharepoint, then Excel can be integrated into Sharepoint and, therefore, be set up quickly as a multi-user environment.

But, it also has some shortcomings. One of Excel's shortcomings is that some view it as a database tool and it is not. One could say that this is a shortcoming but it really is not as it was not designed to be a database.

Excel may have other shortcomings and some of these can be related to technical issues. You can lose data, for example, by importing data into Excel from one format (text) where the original text file has more lines/records than can be handled in the spreadsheet. This can result in lost data records.

Now, experienced and Excel-savvy users will be aware of these limitations and the potential errors the limits can create. However, people can be in a hurry, or your department can be in a crisis-mode (it happens), and limitations might be forgotten or not fully considered. Moreover, it's not easy to know the exact limitations you have at the time as this also can be version-dependent. Even if you discover the limitation in time, some workaround has to be figured out.

Some corrective actions need to be defined - perhaps one of your Excel experts suggests splitting the spreadsheet into several worksheets.  
HmMMM!?

These issues or potential issues mean that you have to be aware and careful when using Excel for "database" type situation.

## What Is A Real Database

What is a "Real Database"? Too often, it seems that Excel is used as a tool to do the work of a "real database". This is often bad idea. Excel is a tool, and just like any tool, it is essential to know what it was designed to do. It is important that the user know its limitations. It is important for managers to know how best to use it.

Imagine you are a manager handling thousands or 10s of thousands of health records. You must use the best tool for the assignment. If you, the manager, are somewhat familiar with Excel you may believe that Excel is an appropriate tool. It's been around for decades. It has been designed and developed by one of the world's top computer technology companies. Excel has a great reputation.

Therefore, it's not surprising that Excel can be assumed to be the right tool to store large numbers of data records in a spreadsheet (database) format. Some may even think that it is a real database as a spreadsheet looks very much like a database to a non-database user.

However, there are tools designed to be actual database tools. Think of systems such as:

- PostgreSQL
- SQL Server
- Oracle
- SQLite

These systems were designed to store large amounts of data. They store and allow access with integrity. They enable specific rules to be defined and set up. For example, you can specify a rule that ensures that data records are not duplicated - that the same data point record is not entered twice.

## Loading A Real Database

Loading a "real database" with data can be daunting as there are so many different ways of importing data. Without experience, setting up the automated loading of a database can be a real challenge. However, Excel and databases can be made to work together. There are efficient ways of setting up automated workflows for these spreadsheet-database scenarios. The best solution is to use the right tools for the right purpose, and Excel can still be a part of that solution.

## An Excel-Database Scenario

Consider this scenario. You regularly collect a set of data from multiple sources (data from numerous offices or departments, for example).

Excel may be the easiest input tool for non-database users who are a valuable and essential part of the data collection process. These users work in the local offices. It may be time-consuming and costly to retrain or replace users that are already fully fully trained and able to work using Excel. Therefore, continue to let them use Excel at the local-office level.

To do this efficiently you could set up an Excel file format using a standard template. Using this Excel template would be the data input tool at the local level. Multiple files would be created locally but all would be in the same format defined by the "universal" template. Each file would have the same format.

These individual Excel files would only contain a portion of the data. When the individual spreadsheet files are combined you will have the complete database. The individual files could contain data inputs by region, or by time frame, or by region and time frame. You would then combine the individual Excel files using a central service. The central service would bring these individual file results together into one single database using a single **ETL** tool.

## Exact Transform Load (ETL)

In database speak, an **extract transform load (ETL)** tool refers to the process of copying data from one or more sources into a destination system-location. The final database represents the data differently from the source(s) or in a different context than the source(s).

## Excel - A Popular & Capable Spreadsheet Tool

The final combined database could be linked so those same local users would still have the ability to produce an overview of all results. This would enable them a means to check their data, collected and input locally, with the final database data.

The final database has been built with input from Excel. Excel was used for department data input. The final database would have been constructed but with key safeguards in place when the Excel data is uploaded to the main database. This would be an acceptable way to combine input from Excel with the final database. Use a database template at the Excel input level to ensure standardization. Then use an **ETL** tool to combine the various Excel spreadsheet inputs with built-in safeguards.

A final single database, with designed-in safeguards helps preserve the integrity of your data.

An ETL tool's automation capabilities, such as **Feature Manipulation Engine (FME)**, can include **Quality Assurance (QA)** rules that send alerts when something suspicious happens. But, it also makes it easier to keep the database up to date. Using **Business Intelligence (BI)** alongside a database is a great way to provide feedback and insights to users.

So how can an ETL tool like FME handle validation and QA in a process like this?

## Your Database - Your Bank Vault

Think of the database as a secure bank-vault. To get data in (and out), it has to have the right credentials. FME is the safety-system checking that your data brings the proper credentials and, if not, warns you that something needs to be investigated.

A few examples of **checks** you could include:

- Are there more (or less) than a certain number of rows in the Excel sheet? This is always a good check as a safety measure.
- Are there duplicate records?
- Are data-values within reasonable ranges?
- Are data-values in the right format (date format versus currency format for example)?
- Do all fields have values?

The list can be much more extensive than this.

One of the most significant benefits of having a well researched and well defined automated process is that you can continuously (if needed) add more checks as the situation evolves! This results in a stronger safety-system as you gain more insights into how well (or not) things are going. Do not blame the tools for misuse of the tool. Blaming Excel for not performing well when it is being used as a final database is not proper. Excel is a spreadsheet. In this capacity it is excellent.

Do not avoid identifying the core problem. Use a tool in the way it was intended to be used. And most importantly, any information owner should take measures to treat their data as a valuable resource.