Engineering data – beyond the spreadsheet

Why engineering with spreadsheets is like building a house with a swiss army knife, and why power tools like Valispace are necessary to do it better.

When explaining the basic concepts of Valispace – collaboration, technical data exchange, connecting data with formulas – a question that often comes up is how Valispace compares with spreadsheets, and especially with cloud-based spreadsheet such as Excel Online and Google Sheets.

It is not a coincidence that spreadsheets are popular: storing data in tables is such a generic approach that it allows for thousands of different applications. Formulas that calculate the value in any particular cell depending on the other cells form a powerful calculation tool that is still relatively easy to learn (although hard to master). Charts can be created easily. Advanced functions like pivot tables and macros allow to set up more elaborate data analysis and automation scenarios.

A spreadsheet is the digital equivalent of a swiss army knife, a flexible tool to 'get stuff done' in our digital lives. Working with spreadsheets has become an important skill to survive and thrive in a digital world, so much so that it forms a mandatory part of secondary school education in many countries.


(https://pixabay.com/en/army-blade-compact-cut-equipment-2185/)

It is not a big stretch to imagine using spreadsheets to set up an engineering data exchange, and in fact this is exactly what is in place in many engineering companies: gather all data in one or multiple big 'master' tables, including the formulas to recalculate cells if other data

changes. Put these on a local shared drive, a document management system or a cloud service such as Excel Online or Google Sheets, to allow for simultaneous editing by different users.

However, large and complex spreadsheets that are in use by multiple people are not a new phenomenon, and the challenges with them have been significant enough to merit quite some academic research and an annual conference. These challenges can include:

- To keep the spreadsheet from becoming a data jungle and allow to retrieve the essential bits of information, typically some **structure** needs to be set up initially and carefully maintained through rules, active cleaning and review.
- The data contains just numbers. For these numbers to be useful in calculations and formulas correction factors are often needed, for example to convert to the correct unit. **Data manipulations** like this need to be clearly **documented** to ensure other users work with the data in the correct way.
- **Changes** are automatically propagated through the formulas, but their effect remains 'hidden' until the correct person happens to looks at the relevant data.

http://dilbert.com/strip/2016-01-06
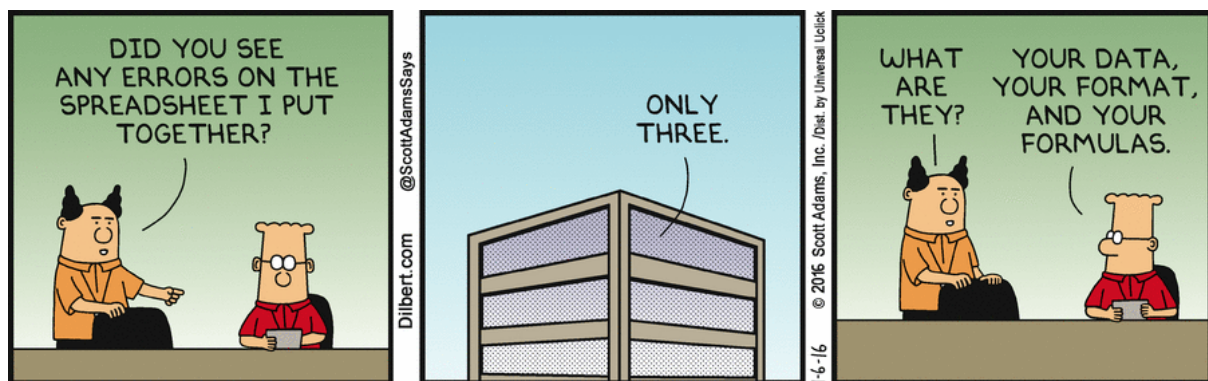
As with any human process, manually maintaining spreadsheets inevitably introduces mistakes. Some of these spreadsheet mistakes have become infamous, but it is increasingly clear that almost every spreadsheet has some bigger or smaller errors. The different types of mistakes in spreadsheets are included in the "Panko-Halverson Taxonomy of Spreadsheet Errors". For example, 'execution errors' can include typos and wrong copies, and 'planning errors' can include logic errors inside formulas.
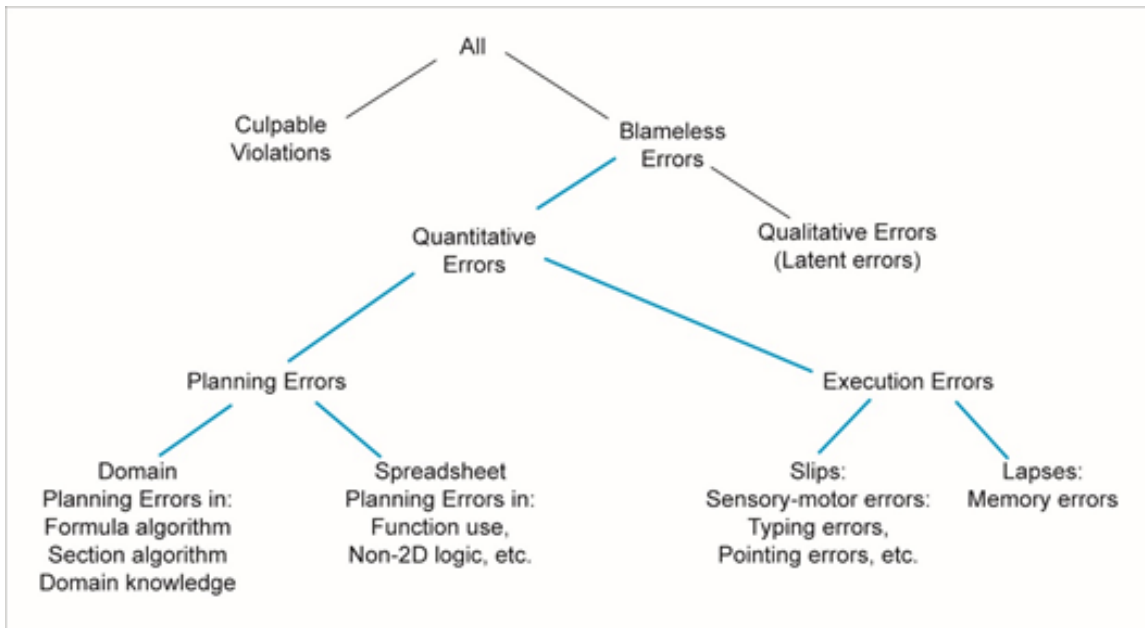
Taxonomy of spreadsheet errors, from [Revising the Panko-Halverson Taxonomy of Spreadsheet Errors](), R. R. Panko and S. Aurigemma, Decision Support Systems 2010.

While the rate of errors in spreadsheets is in itself [not much different]() from human errors in other forms of data storage, reviewing a spreadsheet is notoriously difficult. Have a look at the picture below for a (still relatively straightforward) example. Difficult reviewing is part of the reason why many errors remain hidden and are even copied to other spreadsheets. It is not surprising that these spreadsheet mistakes are often the root cause of expensive last-minute fixes, leading to schedule slips and budget overruns.

| fx	=(G3*VLOOKUP(G2;D1:E10;2)+H3*VLOOKUP(H2;D1:E10;2))/(G3+H3)						
D	**E**	**F**	**G**	**H**	**I**	**J**	
	Price per piece				Average price per piece		
Apple	0,3		Apple	Pear	0,433333		
Pear	0,3	Nb	1	2			
Orange	0,4						
Banana	0,25						
Pineapple	1,2						
Grapefruit	0,7						
Lemon	0,15						
Mango	1,5						
Kiwi	0,5						

Spreadsheet error example. How much time did it take you to find and fix the mistake?

The manual processes of adding structure, performing data manipulations and taking care of changes are introducing errors in spreadsheets that are difficult to find and fix. How can this be solved? It is essential to set up a clear data structure that is shared between all

collaborators, as well as automating as many actions as possible. It is exactly these two solutions that Valispace provides:

1) **Data structure**: any data point inside Valispace is more than just a number. It has a name, a unit, a history, margins and many more. It typically is part of a data 'product tree' (see picture below). Formulas are constructed based on the name of the input values, rather than referring to an abstract cell number, which makes these formulas much easier to review. In every description or calculation where this value is used, it is always referring to the 'single source of truth', ensuring that calculations are consistent among each other.

   At the same time, very rigid data structures can easily make it complex to set up the data storage system and require significant training from all users. That is why a lot of flexibility remains within the basic Valispace 'tree' structure of components, Valis and their properties.
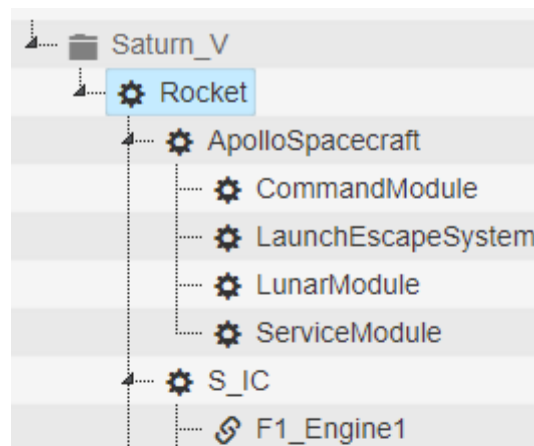
Example component tree from Saturn V advanced tutorial.

2) **Automation**: avoiding manual data manipulations greatly reduces the possibility of human errors. Valispace provides a range of different automation tools:
   a. A lot of important but tedious calculations are performed by the Valispace algorithm behind the scenes, for example: automatic unit conversion, automatic propagation of design margins, notifications to the relevant users when something has changed that they need to be aware of.
   b. Comparison charts, budget tables, relation charts between values and other data summary visuals are automatically to up to date.
   c. Custom automation or scripting can be implemented through the Valispace Python and REST APIs.
   d. Values included in other tools can easily be kept up to date with the Valispace add-ons for Microsoft Word, Excel, MATLAB,…

Through data structure and automation, Valispace solves the key issues of engineering data processing. **While a 'swiss army knife' spreadsheet provides the ultimate flexibility in one tool, more complex data construction work requires dedicated tools to limit the**

amount of manual labor. Valispace is such a tool, a power drill for engineering data that provides efficient data storage and analysis.

https://pixabay.com/en/drilling-machine-drill-power-work-2136837/