Understanding Semantic Spreadsheets

Leveraging XBRL-based, SBRM-based semantic spreadsheets for effective information exchange

By

Charles Hoffman, CPA (Charles.Hoffman@me.com)

Contributors

Andrew Noble, PNA, BBus (Andrew@nobleaccounting.com.au)

Last Revised – August 3, 2020 (DRAFT)

> "The things that we should do are: the things that need to be done, the things you see that need to be done, and the things no one else thinks need to be done." *Richard Buckminster Fuller*[1]

Executive summary:

- This document helps professional accountants understand the notion of a semantic spreadsheet.

[1] WikiQuotes, Richard Buckminster Fuller, https://simple.wikiquote.org/wiki/Richard_Buckminster_Fuller

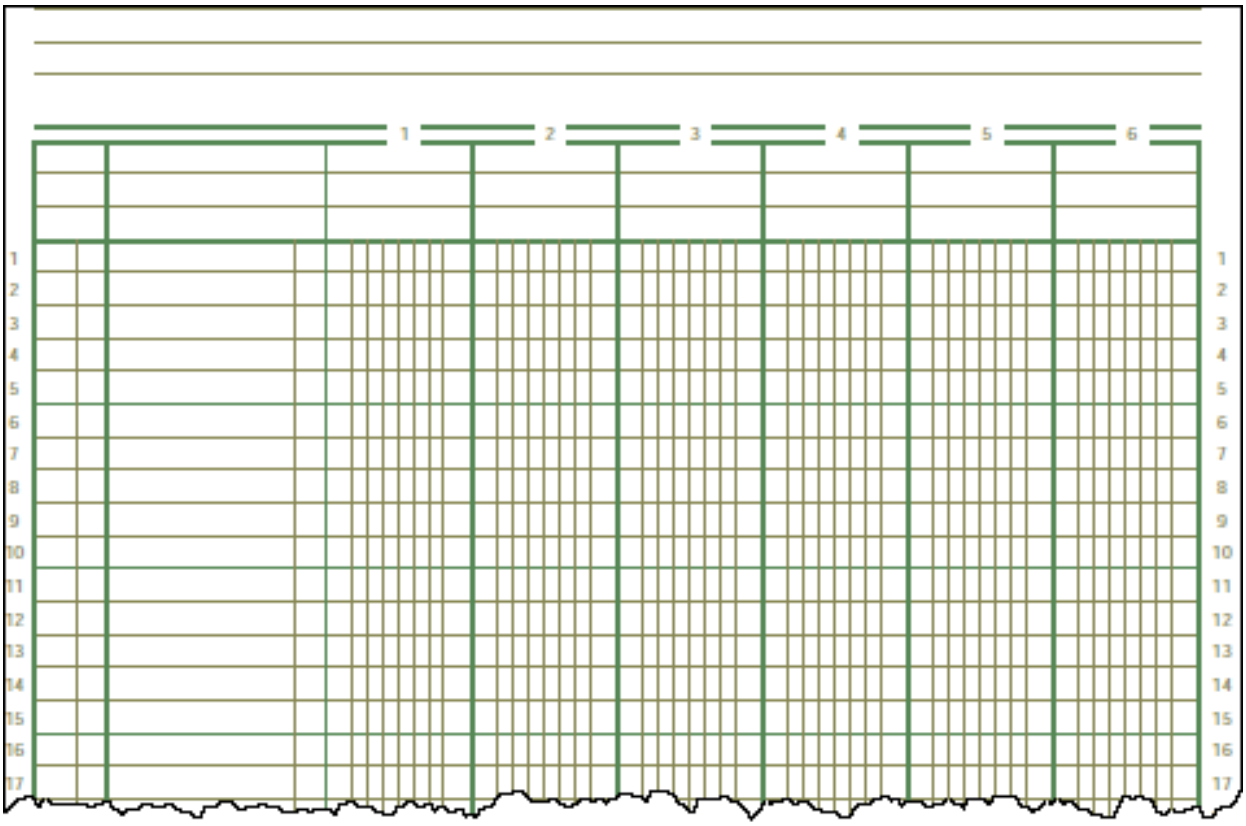
Copyright (full and complete release of copyright)

Acknowledgements:

While I did physically create the information in this resource, I could have not done so without the gracious help of a number of people, directly and indirectly, over the years. I see myself as merely a custodian of this important information, nurturing it along for the benefit of all, condensing countless discussions into something hopefully useful for the common good.

I would like to specifically thank these indirect contributors: Walter Hamscher, PhD, Geoff Shuetrim, PhD; David vun Kannon; Josef Macdonald, CA; Jim Richards; Roger Debreceny; Jeff Naumann, CPA; David Prather, Alan Teixeira, CA; Hugh Wallis; Allyson Ugarte; Colm O hAonghusa; Giancarlo Pellizzari; Yossef Newman, CPA; Rob Blake; Mark Creemers; Marc van Hilvoorde; Herman Fischer; Ignacio Hernandez-Ros; Dean Ritz; Timothy Randle; Cliff Binstock; David Scott Stokes; Masatomo Goto; Paul Warren; Mark Goodhand; Campbell Pryde, CPA; Michele Romanelli; Maciej Piechocki, PhD; Victor Morilla; Mike Rowling; Joe Ryba, CPA; Matthias Brantner; Dennis Knochenwefel; Ghislain Fourny, PhD; Chris Taylor, CPA, Eric Cohen, CPA; Mike Willis, CPA; Louis Matherne, CPA; Hamed Mousavi. There are others which I probably left off and for this I apologize. I acknowledge and appreciate the thinking others contributed to this endeavor.

Historically, a spreadsheet is an accounting ledger that shows what tends to be quantitative information within the rows and columns of a piece of paper and were quite useful in managing a business:



But spreadsheets are not limited to financial data and are frequently used to represent scientific data, engineering information, etc. and commonly used to carry out all sorts of computations.

Killer App

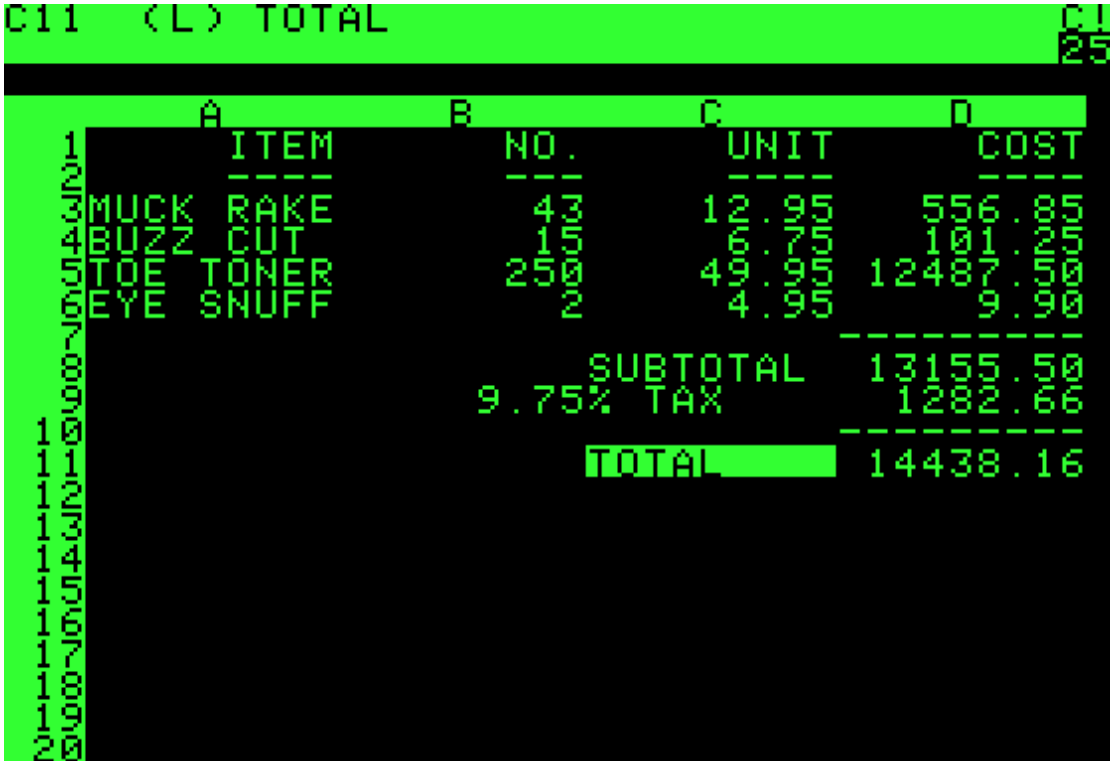
Electronic spreadsheets all but replaced pen-and-ink versions by the end of the 20th century.

In 1979, VisiCalc, which was an easy enough to use electronic spreadsheet that non-nerds could use it and the Apple II went from being a hobbyist's toy to a productive office tool. In 1983, Lotus 1-2-3 was released exclusively for the IBM PC and other MS-DOS computers and it quickly surpassed VisiCalc.

The first known idea for what became the computerized spreadsheet[2] was by Richard Mattessich[3] who pioneered such tools for business accounting at the University of British Columbia, Canada.

In 1969, Rene Pardo and Remy Landau co-invented LANPAR or LANguage for Programming Arrays at Random, an electronic spreadsheet type application[4] that ran on mainframe computers.

But it was not until VisiCalc which was easy enough for the average business professional to use and became the "killer app" that electronic spreadsheets and computers took off.



Today, Microsoft Excel[5] is the undisputed king of the spreadsheet hill. Other spreadsheet tools are commonly used such as Google Sheet[6], Apple Numbers[7], and other spreadsheet applications[8].

[2] Wikipedia, *Spreadsheet*, https://en.wikipedia.org/wiki/Spreadsheet
[3] Wikipedia, *Richard Mattessich*, https://en.wikipedia.org/wiki/Richard_Mattessich
[4] Wikipedia, *Spreadsheet, LANPAR Spreadsheet Compiler*, https://en.wikipedia.org/wiki/Spreadsheet#LANPAR_spreadsheet_compiler
[5] Wikipedia, *Microsoft Excel*, https://en.wikipedia.org/wiki/Microsoft_Excel
[6] Wikipedia, *Google Sheet*, https://en.wikipedia.org/wiki/Google_Sheets
[7] Wikipedia, *Numbers*, https://en.wikipedia.org/wiki/Numbers_(spreadsheet)
[8] Wikipedia, *Comparison of Spreadsheet Software*, https://en.wikipedia.org/wiki/Comparison_of_spreadsheet_software

Two-Dimensional Model

Spreadsheet, computer program that represents information in a tabular[9] or two-dimensional grid of data, along with formulas that relate the data.

My Spreadsheet

	A	B	C	D
01	Sales	100000	30000	70000
02	Purchases	25490	30	200

Multi-Dimensional Model

But information can commonly have more than two dimensions. Information could have three, four, five, or even more useful dimensions. For example, below you see the dimensions "Region", "Gender", and "Style":

	A	B	C	D	E	F	G
1	**Region**	**Gender**	**Style**	**Ship Date**	**Units**	**Price**	**Cost**
2	East	Boy	Tee	1/31/2005	12	11.04	10.42
3	East	Boy	Golf	1/31/2005	12	13	12.6
4	East	Boy	Fancy	1/31/2005	12	11.96	11.74
5	East	Girl	Tee	1/31/2005	10	11.27	10.56
6	East	Girl	Golf	1/31/2005	10	12.12	11.95
7	East	Girl	Fancy	1/31/2005	10	13.74	13.33
8	West	Boy	Tee	1/31/2005	11	11.44	10.94
9	West	Boy	Golf	1/31/2005	11	12.63	11.73
10	West	Boy	Fancy	1/31/2005	11	12.06	11.51
11	West	Girl	Tee	1/31/2005	15	13.42	13.29
12	West	Girl	Golf	1/31/2005	15	11.48	10.67

Pivot Table

A **pivot table**[10] is a table of information that summarizes the information in the form of a more sophisticated table. Information is generally summarized into sums, averages, or other such meaningful ways. Pivot tables are dynamic and literally let users of such models to arrange or rearrange (i.e. "pivot") information to focus on specific information or draw attention to information in specific useful ways.

[9] Wikipedia, *Table (Information)*, https://en.wikipedia.org/wiki/Table_(information)
[10] Wikipedia, Pivot Table, https://en.wikipedia.org/wiki/Pivot_table

Sum of Units	Ship Date					
Region	1/31/2005	2/28/2005	3/31/2005	4/30/2005	5/31/2005	6/30/2005
East	66	80	102	116	127	125
North	96	117	138	151	154	156
South	123	141	157	178	191	202
West	78	97	117	136	150	157
(blank)						
Grand Total	363	435	514	581	622	640

Online Analytical Processing (OLAP)

Online analytical processing, or OLAP[11], is an approach to answering analytical questions using information that is generally taken from an online transaction processing (OLTP)[12] database, put into a data warehouse[13], and then queried using multi-dimensional analytical (MDA) queries. OLAP is part of the broader category of business intelligence applications which also encompasses relational databases, report writing and data mining.

Problems with Spreadsheets

The electronic spreadsheet is the "Swiss army knife" of accountants. While electronic spreadsheets are wonderful tools, electronic spreadsheets are not perfect tools. People point out the flaws of the electronic spreadsheet including[14]:

- Vulnerable to fraud
- Susceptible to trivial human errors
- Difficult to troubleshoot or test
- Obstructive to regulatory compliance
- Unfit for agile business practices
- Not designed for collaborative work
- Hard to consolidate
- Incapable of supporting quick decision making
- Unsuited for business continuity
- Scales poorly

Business professionals tend to love spreadsheets, information technology departments tend to loath electronic spreadsheets for the problems they cause.

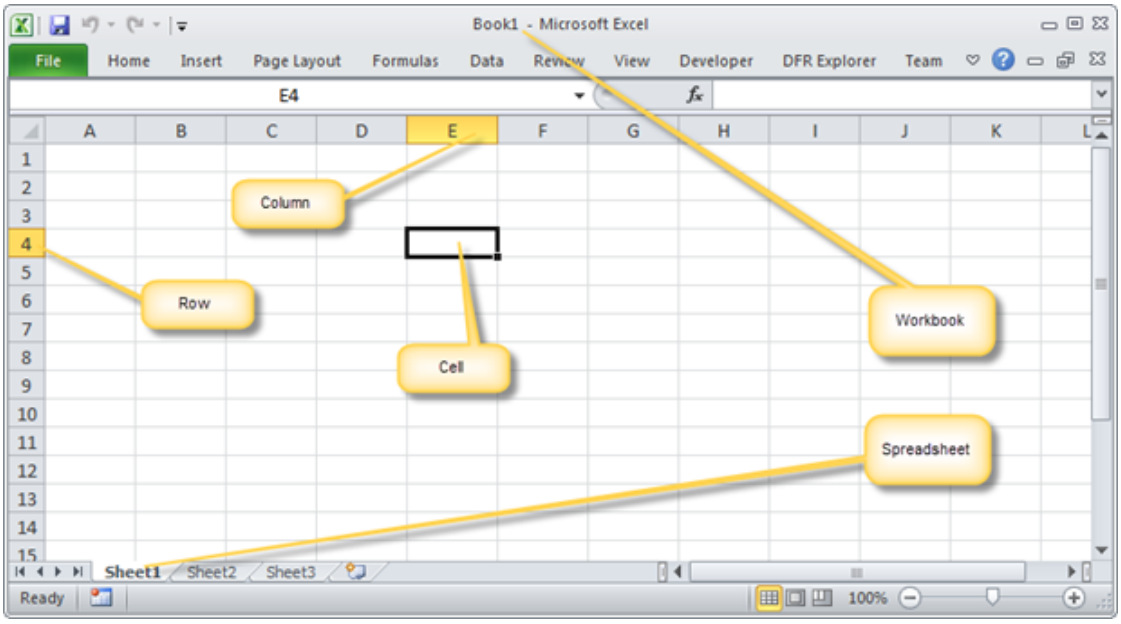
For more information on the problems with spreadsheets, please see Eusprig.org[15].

---

[11] Wikipedia, Online Analytical Processing, https://en.wikipedia.org/wiki/Online_analytical_processing
[12] Wikipedia, Online Transaction Processing, https://en.wikipedia.org/wiki/Online_transaction_processing
[13] Wikipedia, Data Warehouse, https://en.wikipedia.org/wiki/Data_warehouse
[14] Top 10 Disadvantages of Spreadsheets, http://www.denizon.com/spreadsheets/top-10-disadvantages-of-spreadsheets/
[15] Eusprig.org, *European Spreadsheet Risk Interest Group*, http://eusprig.org/

Presentation-based Spreadsheet Model

Another issue with the electronic spreadsheets is that the spreadsheet model is presentation based as contrast to meaning-based model. This is what I mean. The terms “Spreadsheet”, “Row”, “Column”, and “Cell” are all presentation-oriented artifacts:

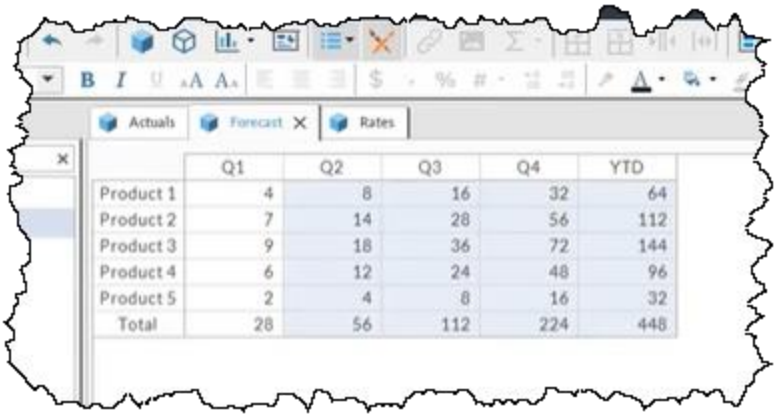


What if we could have a spreadsheet that was meaning-based instead of based on the physical location of the information. For example, rather than referring to current year sales by referring to cell “A1”; we would refer to simply “Sales” and “2020”.

Meaning-based Spreadsheet Model

An alternative to a presentation-based model which uses spreadsheet, row, column, and cell to identify and work with information artifacts is to use a meaning-based spreadsheet model. An example of a meaning-based model can be seen by having a look at the Quantrix modeler[16]. The video is about 11 minutes long and provides a good glimpse into how such a meaning-based spreadsheet might work.

[16] YouTube.com, *Quantrix Introductory Demo | Getting Started with Modeler | Part 1*, https://youtu.be/OE4H29Rz8q8



	Q1	Q2	Q3	Q4	YTD
Product 1	4	8	16	32	64
Product 2	7	14	28	56	112
Product 3	9	18	36	72	144
Product 4	6	12	24	48	96
Product 5	2	4	8	16	32
Total	28	56	112	224	448

Logical Spreadsheets

A logical spreadsheet[17] is a spreadsheet in which formulas take the form of logical constraints rather than function definitions. A similar idea is the notion of a deductive spreadsheet[18]. For more information on logical spreadsheets please see, Introduction to Logical Spreadsheets[19].

Problems with OLAP

OLAP, like any other tool, is not perfect. While some point out micro-level issues with OLAP[20], there are also macro level issues with OLAP. Here is a summary of issues with OLAP[21] that I have accumulated trying to understand and use OLAP tools to make use of XBRL:

- There is no global standard for OLAP
- Cube rigidity, business users cannot change or create new cubes
- Limited computation support, mainly supports only roll ups
- Limited business rule support and inability to exchange business rules between implementations
- Inability to transfer cubes between systems, each system is a "silo" which cannot communicate with other silos
- Inability to articulate metadata which can be shared between OLAP systems
- Focus on numeric-type information and inconsistent support for text data types
- OLAP systems tend to be internally focused within an organization and do not work well externally, for example across a supply chain

[17] Wikipedia, *Logical Spreadsheet*, https://en.wikipedia.org/wiki/Logical_spreadsheet
[18] Deductive Solutions, Iliano Cervesato, *The Deductive Spreadsheet*, https://www.cs.cmu.edu/~iliano/slides/cmu06.pdf
[19] MICHAEL KASSOFF and ANDRE VALENTE, *Introduction to logical spreadsheets*, http://logic.stanford.edu/people/mkassoff/papers/introtologicalspreadsheets.pdf
[20] The Problems with OLAP, http://www.information-management.com/issues/20070301/1076555-1.html
[21] Understanding Cell Stores and NOLAP, the Future of the Spreadsheet, http://xbrl.squarespace.com/journal/2014/11/14/understanding-cell-stores-and-nolap-the-future-of-the-spread.html

- OLAP tends to be read only

Spreadsheet Hell

If you have every used a spreadsheet, they it is highly likely that you have experienced Spreadsheet Hell. Everybody has experienced Spreadsheet Hell at one time or another.

The YouTube video, *The Ununderstood Dark Matter of IT*[22], points out many problems of spreadsheets.

Improving the Spreadsheet

Imagine an improved electronic spreadsheet, a semantic spreadsheet[23] which overcomes many of the problems of how spreadsheets work today. Imagine an improved OLAP, or NOLAP (not only OLAP)[24]; where the spreadsheet is inherently a dynamic pivot table. Imagine a new take on spreadsheets[25]. Imagine an end to what is called "spreadsheet hell"[26].

This is my take on what is wrong with current electronic spreadsheets is this list of 5 fundamental problems that I see and how to fix those problems:

1. **Information is presentation oriented rather than meaning oriented**: Today's electronic spreadsheets, all of them, are made up of sheets which contain rows and columns which intersect to form cells. Information is entered into cells. All these rows, columns, and cells are presentation oriented. What if the information was meaning oriented instead? What if spreadsheet information was glued together by the meaning of the information?
2. **Business rules combined with spreadsheet information**: Spreadsheets today have the data within the spreadsheet combined with the business rules such as formulas for how information adds up, tests to make sure there are no errors, and other information mixed within the data of the spreadsheet. This can make it very hard to check a spreadsheet for errors or missing business rules. To look at this another way, imagine a spreadsheet which is verified using an *external set of business rules*. Sometimes the business rules could be publicly available, other times the business rules would be securely available to a select group of users of the spreadsheet. The basic premise is that you can separate the business rules used to check the spreadsheet from the actual information which provides more control over both the business rules and the information. Plus, this means that the same set of business rules can be used across multiple spreadsheets to verify that the spreadsheets do not contain errors. Considering #1 above, the information, the business rules, and how the information is presented all really

[22] YouTube.com, *The Ununderstood Dark Matter of IT*, https://www.youtube.com/watch?v=wbiVK6HKHHg

[23] See *Semantic spreadsheets*, http://xbrl.squarespace.com/journal/2013/4/18/semantic-spreadsheets.html

[24] See *Understanding Cell Stores and NOLAP, the Future of the Spreadsheet*, http://xbrl.squarespace.com/journal/2014/11/14/understanding-cell-stores-and-nolap-the-future-of-the-spread.html

[25] See *Time for a New Take on the Electronic Spreadsheet*, http://xbrl.squarespace.com/journal/2013/8/2/time-for-a-new-take-on-the-electronic-spreadsheet.html

[26] See *XBRL Ends Spreadsheet Hell*, http://xbrl.squarespace.com/journal/2009/5/2/xbrl-ends-spreadsheet-hell.html

need to be separated to make the spreadsheet more flexible. So, what if business rules could be external to the spreadsheet?

3. **Multiple copies of the same spreadsheet**: A big problem is multiple versions of the same spreadsheet and you lose track of which version is the correct version to be using. Many people refer to this issue as spreadsheet hell. More and more people are addressing this by storing spreadsheet information in a database and exposing the information view Excel, but saving the information into a database. The problem with this is see #1 above, the information stored is still presentation oriented and not meaning oriented. What if you addressed information by the meaning of the information, the characteristics of the information is how you identify the information?[27]
4. **Comparing information between spreadsheets can be a challenge**: If you have ever given a spreadsheet to two or more different people, had each person put information into the spreadsheet, and then tried to compare spreadsheet information you understand this situation. Reusing information contained in spreadsheets effectively can be a big challenge. What if you could compare meaning?
5. **Proprietary format, forced to use one software application**: Excel is a great software application for working with spreadsheets. But if you don't have Excel or someone you want to share information with does not have Excel and you want to exchange information, this can be problematic. The interoperability between Excel, Google Spreadsheets, and Apple Numbers spreadsheets is OK some times, but other times problematic. Standard formats such as Open Documents helps, but the standards focus on formatting of information, not the semantics of the information. Also, business rules are still embedded within the application. Further, Excel is a very "heavy" client. With tablet PCs and mobile devices growing in popularity, that becomes more and more of a problem. What if a spreadsheet was a global standard format, rather than a proprietary format of one software vendor?

Requirements for an Improved Electronic Spreadsheet

Here are my requirements for a better spreadsheet[28]. This new improved version of a spreadsheet is not intended to replace 100% of all existing spreadsheets. Rather, this is intended to be a new alternative, a new category of spreadsheet. An alternative which could be used in 20% of the cases where more control is needed over spreadsheets (but I suspect the spreadsheet would be use in 80% of cases).

1. **Readable by both humans and machines**: A spreadsheet should be readable by both humans and machines. Information provided within a spreadsheet should be more a representation of information than presentation oriented. The representation can be presented in sheets, rows, columns, and cells but this is done leveraging information metadata and commonly understood patterns. 100% pixel perfect renderings are specifically not a requirement.

[27] One software vendor calls this a "cell store", I believe a better term is a "fact store".

[28] See, *Need for New Global Standard Spreadsheet Alternative*, http://xbrl.squarespace.com/journal/2014/5/3/need-for-new-global-standard-spreadsheet-alternative.html

2. **Global standard format**: The format of the spreadsheet should be a global standard, not controlled by one software vendor.
3. **Agreed upon level of semantics**: The creators and consumers can agree on the level of semantic clarity they will make use of for a spreadsheet. The spectrum can range from no semantics at all (which is similar to today's spreadsheet) or a high level of semantics expressed by a highly controlled representation model.
4. **Separation of representation and presentation**: The "representation" and the "presentation of the representation" should not be intermingled.
5. **Business rules separable from spreadsheet**: Business rules should be separated from the information when desired, integrated with the spreadsheet when necessary. Business rules which are external to the spreadsheet can be used to "watch over" the things and relations within the spreadsheet. The business rules can be made available publicly via a URL, privately via a private URL, etc.
6. **Managed global standard**: The better spreadsheet should be a global standard under the control of someone like OMG, XBRL International, ISO, Apache OpenOffice, or some other such organization.
7. **Provide a formal shape but be domain neutral**: One formal shape should be agreed to, for example the multidimensional model, but the pieces which fit into that shape or "fill" the shape are domain neutral, controlled by the business domain.
8. **Format should allow for versioning, collaboration, etc.**: The syntax format should allow for ease of versioning, constructing systems which are collaborative in nature (multi-user).
9. **Straightforwardly usable over the Internet**: The format should be compliant with internet standards.
10. **Support a wide variety of common business use cases**: A wide variety of common business use cases would be served, but it is not a goal to solve every business problem which exists.
11. **Highly limited options**: The number of optional features is to be kept to the absolute minimum, ideally zero. Multiple approaches to solving a problem are not necessary when one will do.
12. **Formal and concise design**: The design must be formal, concise, well designed and well-engineered.

That is my list. What might you include in your list?

OMG Semantic Spreadsheet

There is the beginnings of an effort at OMG to create what they call a *Semantic Workbook*[29]. That semantic workbook standard would very likely be based on the OMG *Standard Business Report Model* (SBRM)[30]. One technical format for the standard semantic workbook would very likely be XBRL.

[29] OMG Semantic Workbook (SXLM), http://xbrl.squarespace.com/journal/2020/1/16/omg-semantic-workbook-sxlm.html
[30] SBRM Progress Report, http://xbrl.squarespace.com/journal/2020/1/30/sbrm-progress-report.html

The table below reconciles XBRL technical syntax terms, SBRM terms, proposed semantic workbook terms, and W3C's *RDF Data Cube Vocabulary*[31] terms:

XBRL Syntax Terminology (XBRL)	Logical Conceptulization of Business Report (SBRM)	OMG Semantic Workbook (SXLM)	RDF Data Cube Vocabulary
XBRL instance + XBRL taxonomy	Report	Workbook	Data Set
Network + Hypercube	Structure	Worksheet; Pivot Table	Data Structure Definition
XBRL linkbases	Aspect, Association	Row, Column, Filter (Slicer)	Dimensions, Attributes, Measures, Slices
XBRL formula; XBRL definition relations; XBRL calculation relations; XBRL presentation relations	Rule	Rule	Rule
Fact	Fact	Cell	Observation
XBRL Footnote	Parenthetical Explanation	Cell Comment	Label; Comment;
Report Package	Report Set	Set of Workbooks	

An example implementation of a semantic spreadsheet can be found here, implemented using XBRL: http://xbrlsite.azurewebsites.net/2020/master/spreadsheet/

Additional Information

This is additional information on this topic:

https://www.cs.cmu.edu/~iliano/slides/cmu06.pdf

https://www3.cs.stonybrook.edu/~cram/Papers/RRW_KER07/paper.pdf

https://en.wikipedia.org/wiki/Logical_spreadsheet

https://www.cs.cmu.edu/~iliano/papers/ker06.pdf

https://link.springer.com/book/10.1007/978-3-642-37747-1

https://www.kobo.com/us/en/ebook/the-deductive-spreadsheet

https://patents.google.com/patent/US7761782B1/en

https://slideplayer.com/slide/9325019/

[31] W3C, RDF Cube Vocabulary, https://www.w3.org/TR/vocab-data-cube/