ONCE UPON A FILE

FUNNY STORIES ABOUT FRUSTRATING EXPERIENCES WITH FILES AND PEOPLE

SimCorp Sofia

Leading provider of integrated investment management solutions for the Italian insurance market since 1983

More than 50 institutional investors

At the beginning: Italian clients
(using Italian as official language)

Now: international clients mostly located in Europe
(using English as official language)

When working with software is pretty much impossible not to have to deal with:

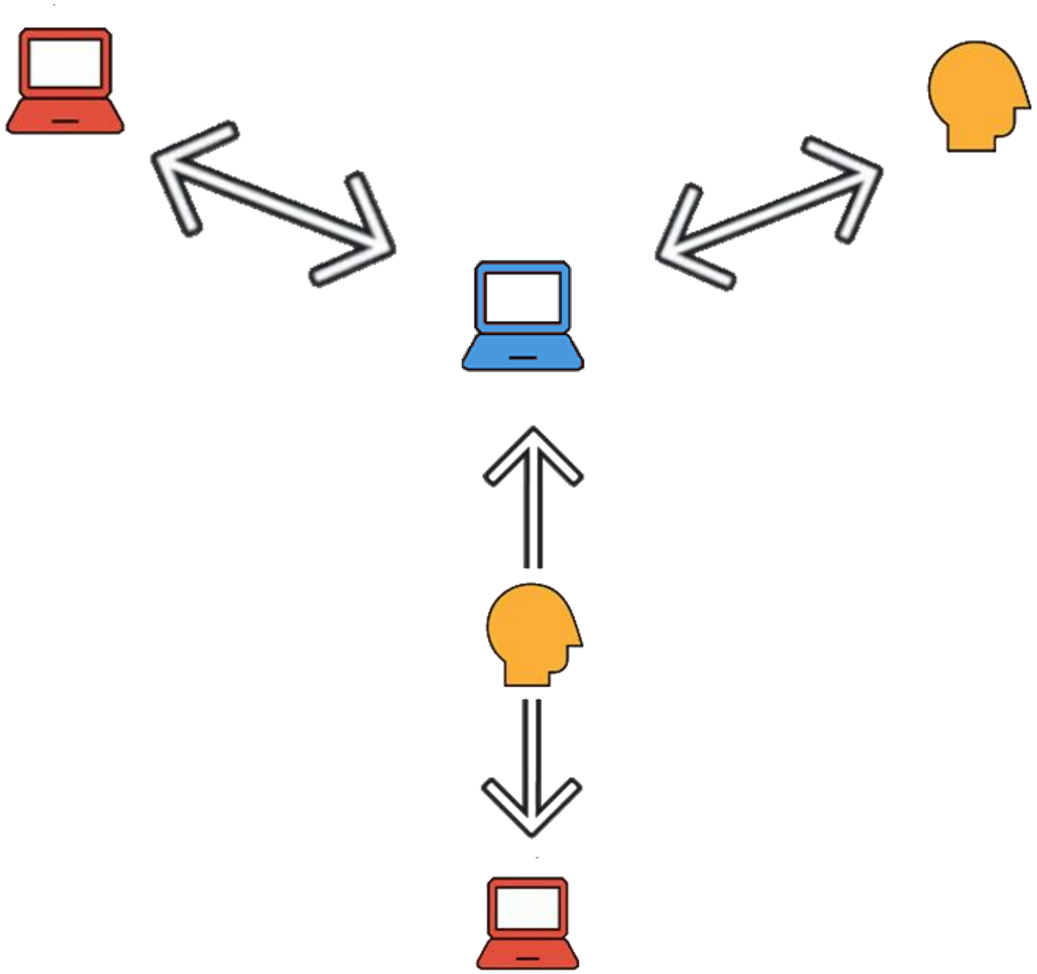
- imports, to get data into your application
- exports, to send data to users or other applications in the outside world

They can sometimes fail and this leads to frustrating experiences…

… that can become funny anecdotes after they are solved!

⚠ All examples in this presentation are based on true stories ⚠

3 types of interaction



- Software vs software
- Software vs human
- Software vs software with human check

The types of problems that can occur are also influenced by

- File format
- File encoding
- Ambiguity of some fields (e.g. dates)

What can happen when the interaction goes wrong?

- Best-case scenario: system error (at least we know we have a problem)

- Not-so-good-case scenario: the import/export works, but the data is spoilt!

In any case, we have to understand where the problem originates and fix it, if it comes from the application, or collaborate with the user, if it comes from the file

Most used import/export formats in Sofia

- Text file with fixed width in ANSI encoding
- Text file with fixed width and delimiter in ANSI encoding
- CSV
- XML
- Excel

A common issue with plain text files

It’s important to know whether the file was produced by Windows or Linux/Unix or Mac OS to determine the end of line char (\r\n, \n or \r)

In theory, this information is valid forever,
as long as you are using the same file from
the same source…

But then…

A user opens the file in a different OS and saves it: the end of line char changes and suddenly the import doesn’t work anymore

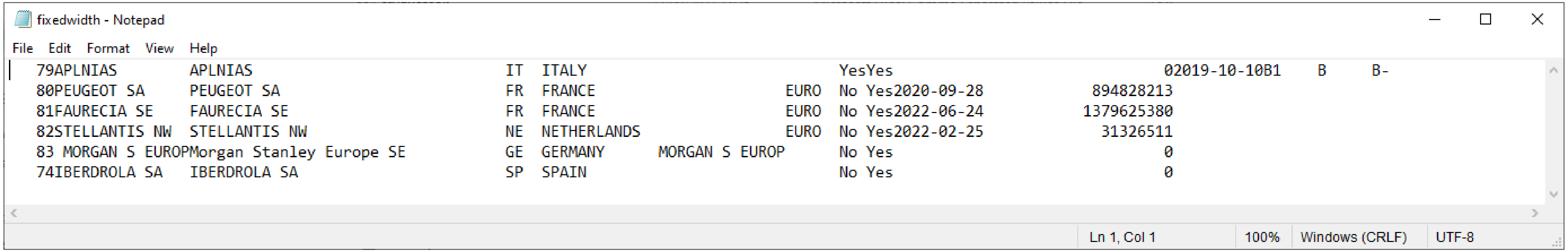
Possible solutions:

Ask the user to import the original file

If they need to modify the original file, they can do it in the original OS

If they can’t use the original OS, some text editors allow to choose the end of line char

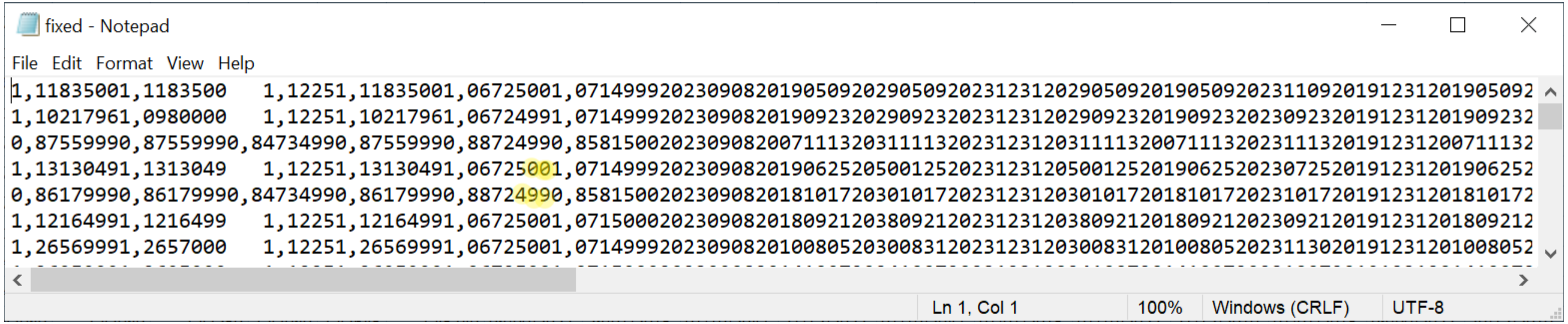
Fixed width



```
   79APLNIAS        APLNIAS                            IT  ITALY                            YesYes                              02019-10-10B1    B     B-
   80PEUGEOT SA     PEUGEOT SA                         FR  FRANCE                     EURO  No Yes2020-09-28            894828213
   81FAURECIA SE    FAURECIA SE                        FR  FRANCE                     EURO  No Yes2022-06-24           1379625380
   82STELLANTIS NW  STELLANTIS NW                      NE  NETHERLANDS                EURO  No Yes2022-02-25             31326511
   83 MORGAN S EUROPMorgan Stanley Europe SE           GE  GERMANY      MORGAN S EUROP      No Yes                              0
   74IBERDROLA SA   IBERDROLA SA                       SP  SPAIN                            No Yes                              0
```

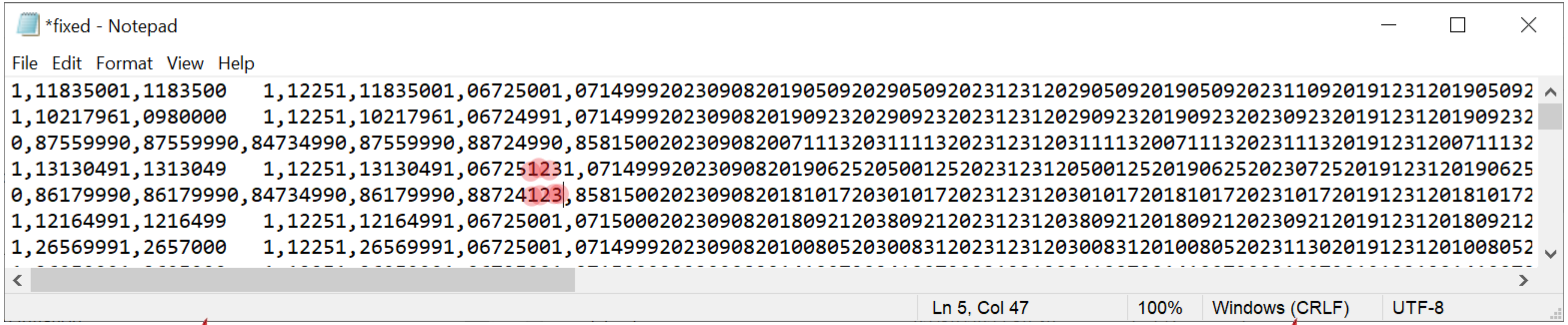
- 🙂 Easy to read for a machine!
- 😐 Managing the fixed width format requires having a list of field widths describing the file
- 🙁 There are a lot of useless blanks

User readibility



```
1,11835001,1183500   1,12251,11835001,06725001,07149992023090820190509202905092023123120290509201905092023110920191231201905092
1,10217961,0980000   1,12251,10217961,06724991,07149992023090820190923202909232023123120290923201909232023092320191231201909232
0,87559990,87559990,84734990,87559990,88724990,85815002023090820071113203111132023123120311113200711132023111320191231200711132
1,13130491,1313049   1,12251,13130491,06725001,07149992023090820190625205001252023123120500125201906252023072520191231201906252
0,86179990,86179990,84734990,86179990,88724990,85815002023090820181017203010172023123120301017201810172023101720191231201810172
1,12164991,1216499   1,12251,12164991,06725001,07150002023090820180921203809212023123120380921201809212023092120191231201809212
1,26569991,2657000   1,12251,26569991,06725001,07149992023090820100805203008312023123120300831201008052023113020191231201008052
```

Editing is difficult: wrong width/wrong data



```
1,11835001,1183500   1,12251,11835001,06725001,07149992023090820190509202905092023123120290509201905092023110920191231201905092
1,10217961,0980000   1,12251,10217961,06724991,07149992023090820190923202909232023123120290923201909232023092320191231201909232
0,87559990,87559990,84734990,87559990,88724990,85815002023090820071113203111132023123120311113200711132023111320191231200711132
1,13130491,1313049   1,12251,13130491,067251231,0714999202309082019062520500125202312312050012520190625202307252019123120190625
0,86179990,86179990,84734990,86179990,88724123,85815002023090820181017203010172023123120301017201810172023101720191231201810172
1,12164991,1216499   1,12251,12164991,06725001,07150002023090820180921203809212023123120380921201809212023092120191231201809212
1,26569991,2657000   1,12251,26569991,06725001,07149992023090820100805203008312023123120300831201008052023113020191231201008052
```

Import crashes or, even worse, data is wrong!

In Sofia, the fixed width text file comes with ANSI encoding

One char is one byte!

When dealing with ANSI file with fixed widths:

- The number of bytes in one line is known beforehand
- It is easy to read big files in chunks parsing always complete lines
- It is easy to identify a certain field in a certain record, just calculate the offset bytes!

ANSI encoding had been enough as long as we had only

Italian customers using Italian language

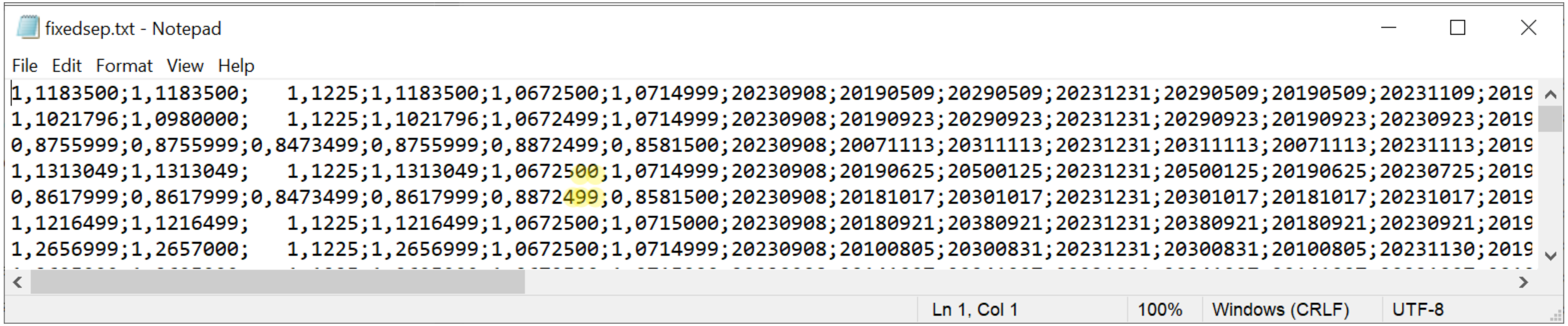
Therefore, we enjoyed the benefit of a 1 byte =

1 char encoding for a while

Fixed width with delimiter

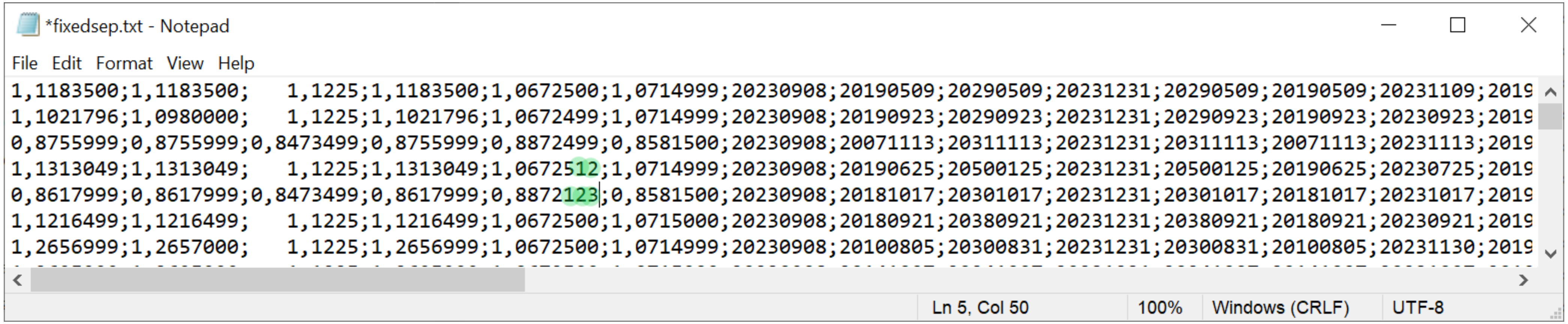
- It is a mix of fixed width and CSV file
- Easier to read for the user, while a program can process it either as a CSV or as a fixed width file
- It has extra blank spaces (so the file is bigger than a CSV) and it doesn’t support the delimiter as a character inside a field (as a CSV does)

Easy to check!　　Even with «difficult» data



```
1,1183500;1,1183500;   1,1225;1,1183500;1,0672500;1,0714999;20230908;20190509;20290509;20231231;20290509;20190509;20231109;2019
1,1021796;1,0980000;   1,1225;1,1021796;1,0672499;1,0714999;20230908;20190923;20290923;20231231;20290923;20190923;20230923;2019
0,8755999;0,8755999;0,8473499;0,8755999;0,8872499;0,8581500;20230908;20071113;20311113;20231231;20311113;20071113;20231113;2019
1,1313049;1,1313049;   1,1225;1,1313049;1,0672500;1,0714999;20230908;20190625;20500125;20231231;20500125;20190625;20230725;2019
0,8617999;0,8617999;0,8473499;0,8617999;0,8872499;0,8581500;20230908;20181017;20301017;20231231;20301017;20181017;20231017;2019
1,1216499;1,1216499;   1,1225;1,1216499;1,0672500;1,0715000;20230908;20180921;20380921;20231231;20380921;20180921;20230921;2019
1,2656999;1,2657000;   1,1225;1,2656999;1,0672500;1,0714999;20230908;20100805;20300831;20231231;20300831;20100805;20231130;2019
```

Editing still requires attention, but the delimiter helps



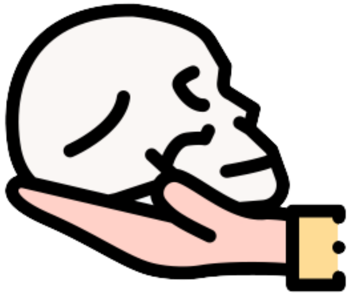
```
1,1183500;1,1183500;   1,1225;1,1183500;1,0672500;1,0714999;20230908;20190509;20290509;20231231;20290509;20190509;20231109;2019
1,1021796;1,0980000;   1,1225;1,1021796;1,0672499;1,0714999;20230908;20190923;20290923;20231231;20290923;20190923;20230923;2019
0,8755999;0,8755999;0,8473499;0,8755999;0,8872499;0,8581500;20230908;20071113;20311113;20231231;20311113;20071113;20231113;2019
1,1313049;1,1313049;   1,1225;1,1313049;1,0672512;1,0714999;20230908;20190625;20500125;20231231;20500125;20190625;20230725;2019
0,8617999;0,8617999;0,8473499;0,8617999;0,8872123;0,8581500;20230908;20181017;20301017;20231231;20301017;20181017;20231017;2019
1,1216499;1,1216499;   1,1225;1,1216499;1,0672500;1,0715000;20230908;20180921;20380921;20231231;20380921;20180921;20230921;2019
1,2656999;1,2657000;   1,1225;1,2656999;1,0672500;1,0714999;20230908;20100805;20300831;20231231;20300831;20100805;20231130;2019
```

CSV

🙂 It gets rid of useless spaces

🙂 It allows to quote fields, so that the delimiter can be used as a character inside other fields

«Quotation marks» or not «Quotation marks»

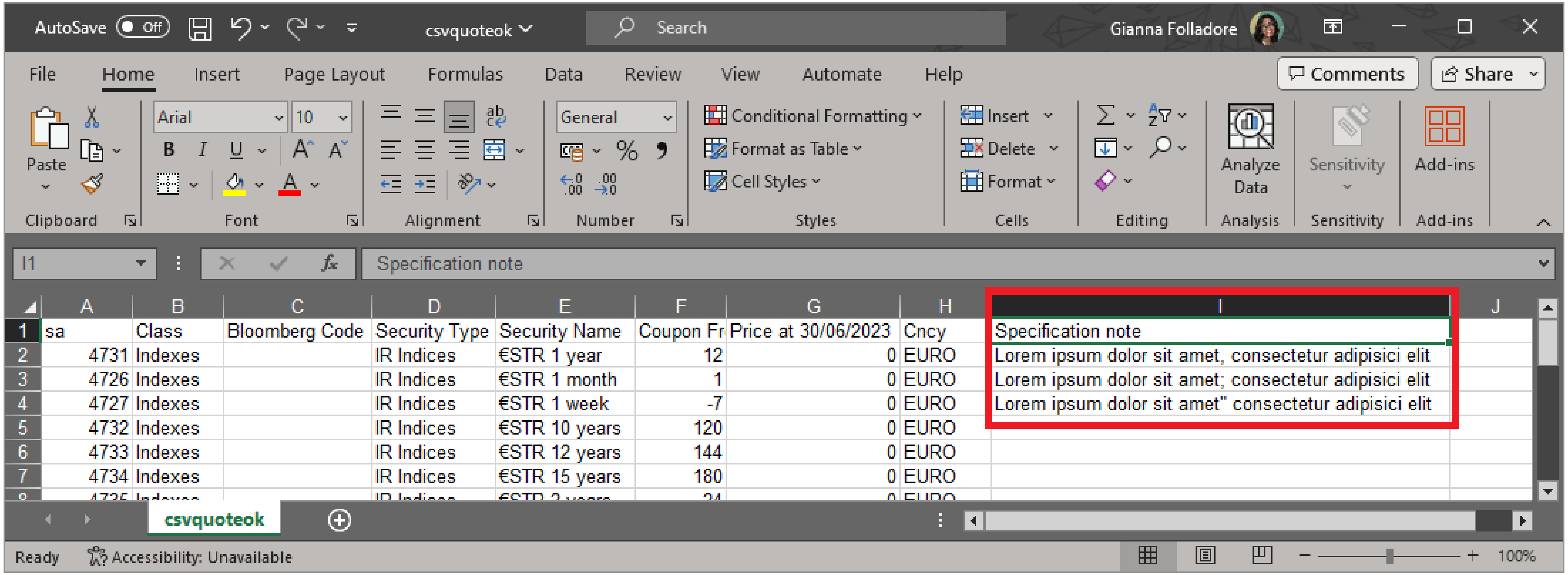


☺ Mandatory if the field includes the text qualifier or the delimiter

☺ An option in all other situations involving text fields

☹ In our experience, it is not always easy to make a user understand how or when use them

This is the information we want to get from a CSV file

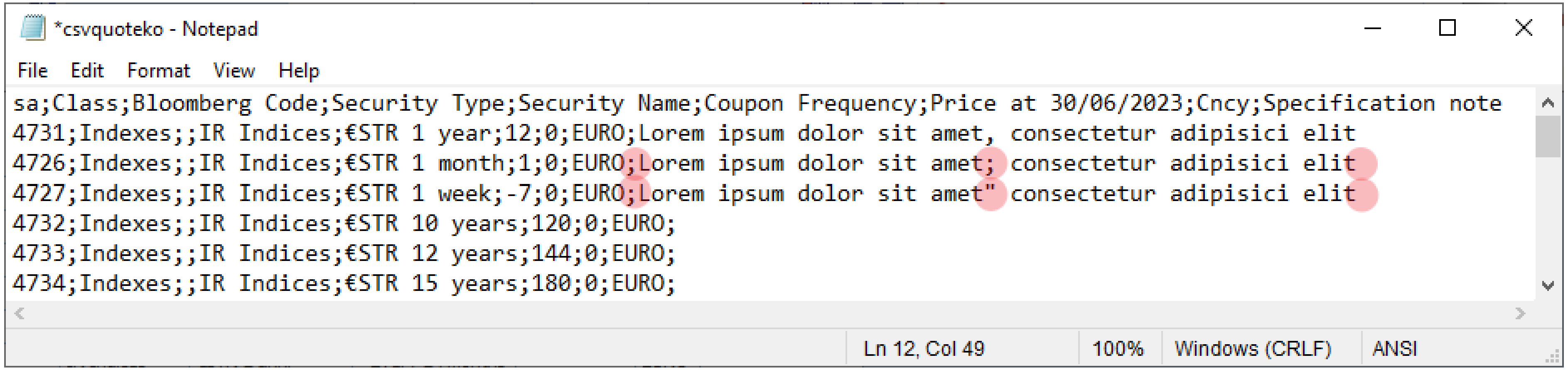


I1 | Specification note

	A	B	C	D	E	F	G	H	I
1	sa	Class	Bloomberg Code	Security Type	Security Name	Coupon Fr	Price at 30/06/2023	Cncy	Specification note
2	4731	Indexes		IR Indices	€STR 1 year	12	0	EURO	Lorem ipsum dolor sit amet, consectetur adipisici elit
3	4726	Indexes		IR Indices	€STR 1 month	1	0	EURO	Lorem ipsum dolor sit amet; consectetur adipisici elit
4	4727	Indexes		IR Indices	€STR 1 week	-7	0	EURO	Lorem ipsum dolor sit amet" consectetur adipisici elit
5	4732	Indexes		IR Indices	€STR 10 years	120	0	EURO	
6	4733	Indexes		IR Indices	€STR 12 years	144	0	EURO	
7	4734	Indexes		IR Indices	€STR 15 years	180	0	EURO	
8	4735	Indexes		IR Indices	€STR 2 years	24	0	EURO	

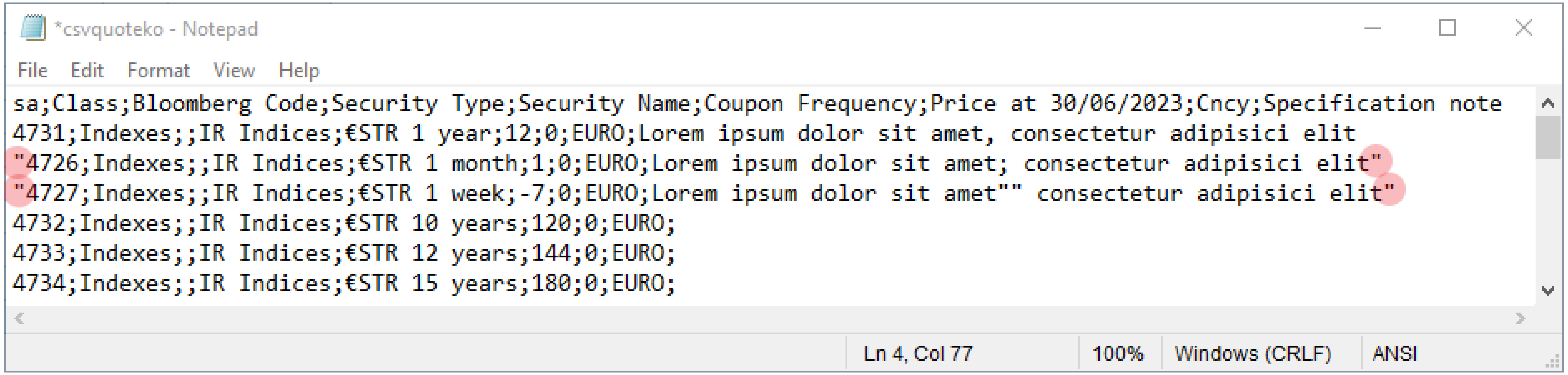
In this example, the semicolon is the delimiter
and the double quote the text qualifier

Second line has one field too many, third line has an invalid field format



```
sa;Class;Bloomberg Code;Security Type;Security Name;Coupon Frequency;Price at 30/06/2023;Cncy;Specification note
4731;Indexes;;IR Indices;€STR 1 year;12;0;EURO;Lorem ipsum dolor sit amet, consectetur adipisici elit
4726;Indexes;;IR Indices;€STR 1 month;1;0;EURO;Lorem ipsum dolor sit amet; consectetur adipisici elit
4727;Indexes;;IR Indices;€STR 1 week;-7;0;EURO;Lorem ipsum dolor sit amet" consectetur adipisici elit
4732;Indexes;;IR Indices;€STR 10 years;120;0;EURO;
4733;Indexes;;IR Indices;€STR 12 years;144;0;EURO;
4734;Indexes;;IR Indices;€STR 15 years;180;0;EURO;
```

Solution: «Just add quotation marks!»

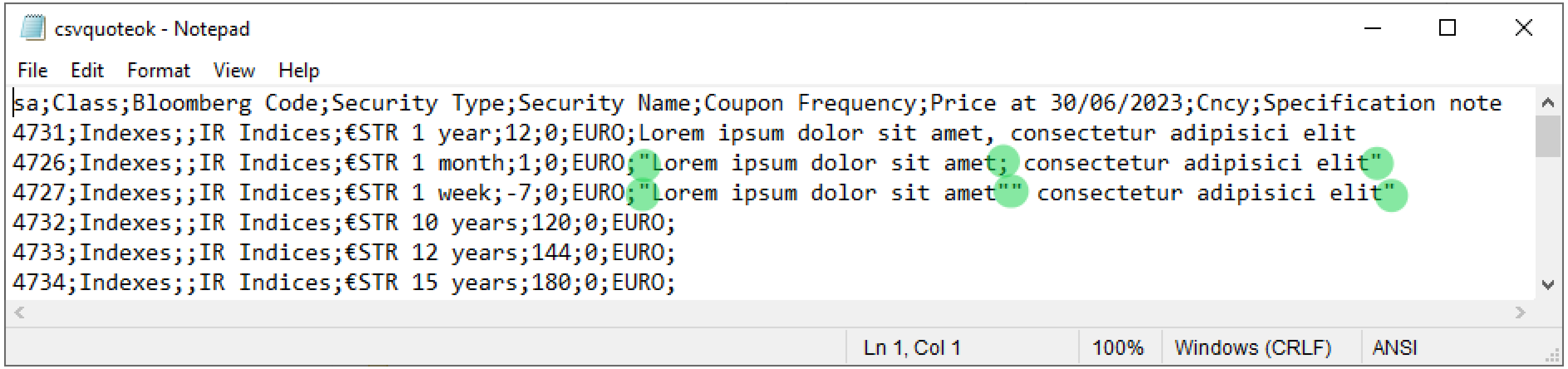


```
sa;Class;Bloomberg Code;Security Type;Security Name;Coupon Frequency;Price at 30/06/2023;Cncy;Specification note
4731;Indexes;;IR Indices;€STR 1 year;12;0;EURO;Lorem ipsum dolor sit amet, consectetur adipisici elit
"4726;Indexes;;IR Indices;€STR 1 month;1;0;EURO;Lorem ipsum dolor sit amet; consectetur adipisici elit"
"4727;Indexes;;IR Indices;€STR 1 week;-7;0;EURO;Lorem ipsum dolor sit amet"" consectetur adipisici elit"
4732;Indexes;;IR Indices;€STR 10 years;120;0;EURO;
4733;Indexes;;IR Indices;€STR 12 years;144;0;EURO;
4734;Indexes;;IR Indices;€STR 15 years;180;0;EURO;
```

The quoted lines are just one big field now!

Solution: «Just add quotation marks at the beginning and at the end of the fields that need them!»

This is a good file



```
sa;Class;Bloomberg Code;Security Type;Security Name;Coupon Frequency;Price at 30/06/2023;Cncy;Specification note
4731;Indexes;;IR Indices;€STR 1 year;12;0;EURO;Lorem ipsum dolor sit amet, consectetur adipisici elit
4726;Indexes;;IR Indices;€STR 1 month;1;0;EURO;"Lorem ipsum dolor sit amet; consectetur adipisici elit"
4727;Indexes;;IR Indices;€STR 1 week;-7;0;EURO;"Lorem ipsum dolor sit amet"" consectetur adipisici elit"
4732;Indexes;;IR Indices;€STR 10 years;120;0;EURO;
4733;Indexes;;IR Indices;€STR 12 years;144;0;EURO;
4734;Indexes;;IR Indices;€STR 15 years;180;0;EURO;
```

Even exports can bring some surprises

Sometimes we need to interact with programs that not fully support the CSV format.
In some cases, the program couldn’t manage the fact that some records had a quoted field, and some didn’t

Possible solutions:
Add quotation marks to all text fields, even when it is not needed
Remove text qualifier and delimiter from text fields

Are CSV files easy to read and edit for the users?

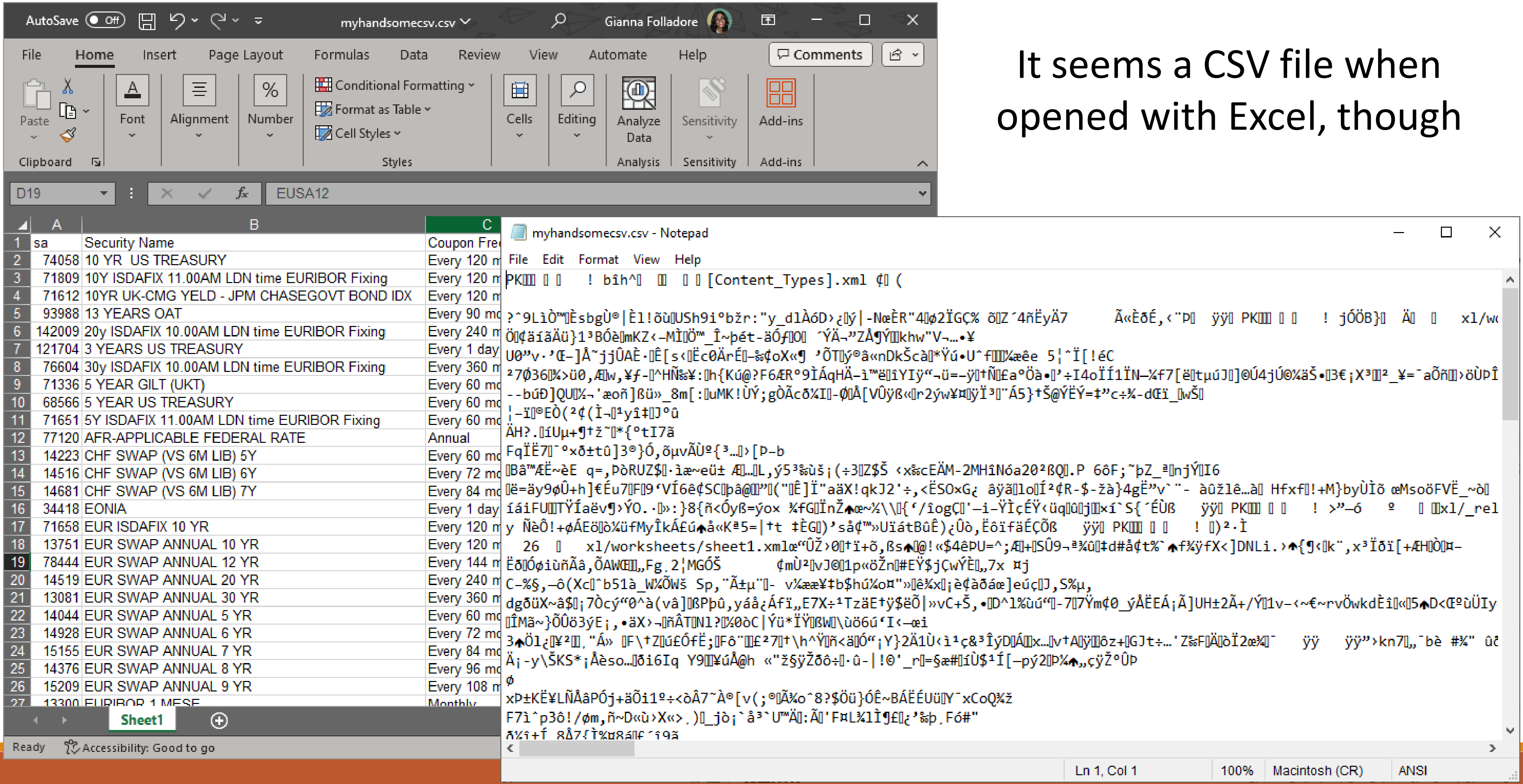
By default, on Windows, CSV files are opened using Excel

This can stir up some confusion between CSV files and Excel files

The results are amazing…

Putting «csv» as extension of an Excel file doesn’t do the trick: this is not a CSV file!!!

It seems a CSV file when opened with Excel, though

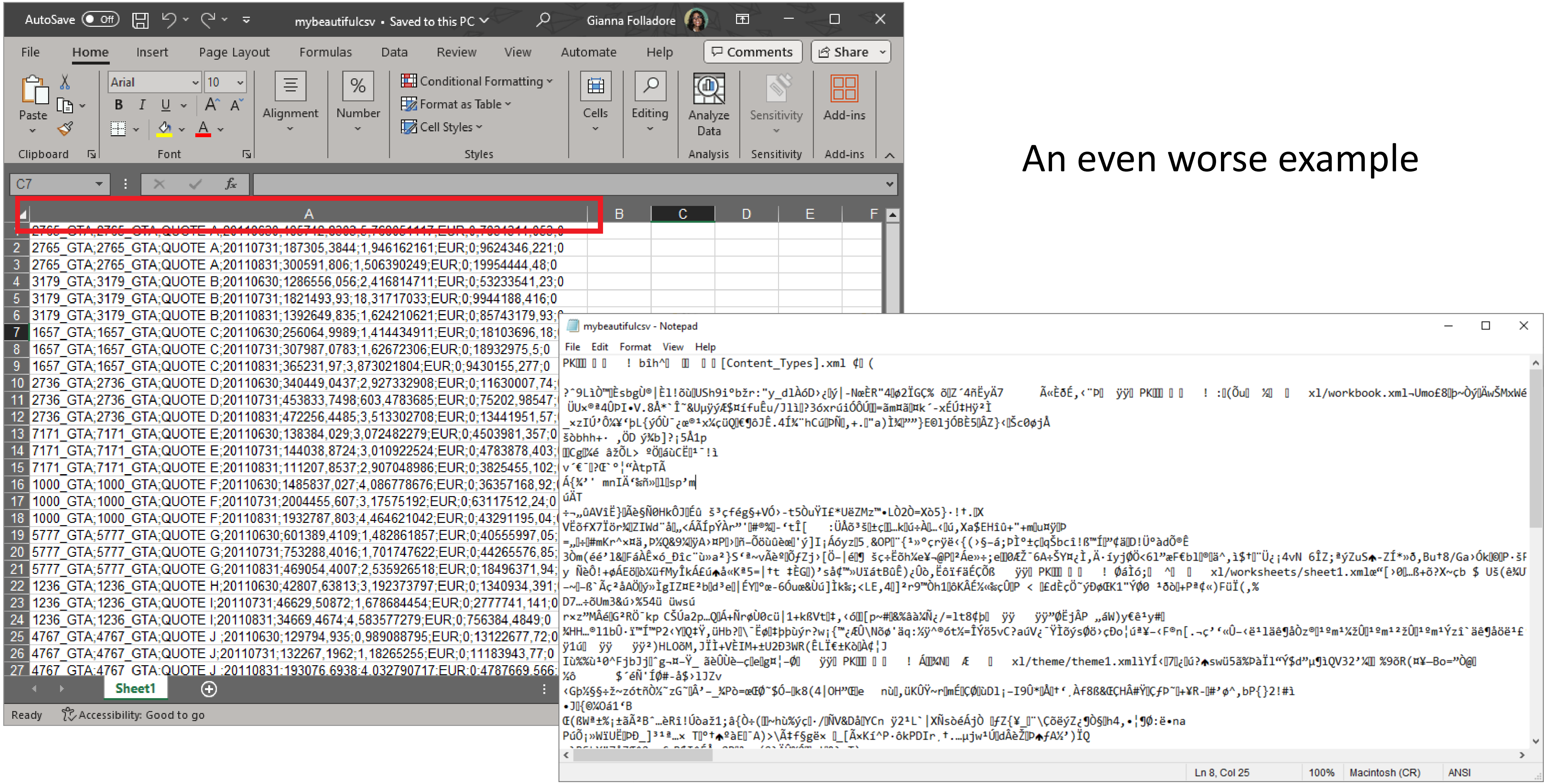


D19 EUSA12

	A	B	C
1	sa	Security Name	Coupon Fre
2	74058	10 YR US TREASURY	Every 120 m
3	71809	10Y ISDAFIX 11.00AM LDN time EURIBOR Fixing	Every 120 m
4	71612	10YR UK-CMG YELD - JPM CHASEGOVT BOND IDX	Every 120 m
5	93988	13 YEARS OAT	Every 90 mo
6	142009	20y ISDAFIX 10.00AM LDN time EURIBOR Fixing	Every 240 m
7	121704	3 YEARS US TREASURY	Every 1 day
8	76604	30y ISDAFIX 10.00AM LDN time EURIBOR Fixing	Every 360 m
9	71336	5 YEAR GILT (UKT)	Every 60 mo
10	68566	5 YEAR US TREASURY	Every 60 mo
11	71651	5Y ISDAFIX 11.00AM LDN time EURIBOR Fixing	Every 60 mo
12	77120	AFR-APPLICABLE FEDERAL RATE	Annual
13	14223	CHF SWAP (VS 6M LIB) 5Y	Every 60 mo
14	14516	CHF SWAP (VS 6M LIB) 6Y	Every 72 mo
15	14681	CHF SWAP (VS 6M LIB) 7Y	Every 84 mo
16	34418	EONIA	Every 1 day
17	71658	EUR ISDAFIX 10 YR	Every 120 m
18	13751	EUR SWAP ANNUAL 10 YR	Every 120 m
19	78444	EUR SWAP ANNUAL 12 YR	Every 144 m
20	14519	EUR SWAP ANNUAL 20 YR	Every 240 m
21	13081	EUR SWAP ANNUAL 30 YR	Every 360 m
22	14044	EUR SWAP ANNUAL 5 YR	Every 60 mo
23	14928	EUR SWAP ANNUAL 6 YR	Every 72 mo
24	15155	EUR SWAP ANNUAL 7 YR	Every 84 mo
25	14376	EUR SWAP ANNUAL 8 YR	Every 96 mo
26	15209	EUR SWAP ANNUAL 9 YR	Every 108 m
27	13300	EURIBOR 1 MESE	Monthly

PK ! bîh^ [Content_Types].xml ¢ (
?ˆ9LìÒ™ÈsbgÙº|Èl!õùUSh9iºbžr:"y_dlÀóD›¿ý|-NœÈR"4ø2ÏGÇ% õZ´4ñËyÄ7 Ã«ÈõÉ,‹"Þ ÿÿ PK ! jÕÖB} Ä xl/w
...
26 xl/worksheets/sheet1.xml
...

An even worse example



	A
1	2765_GTA;2765_GTA;QUOTE A;20110630;185742,8303;5,769051117;EUR;0;7834311,853;0
2	2765_GTA;2765_GTA;QUOTE A;20110731;187305,3844;1,946162161;EUR;0;9624346,221;0
3	2765_GTA;2765_GTA;QUOTE A;20110831;300591,806;1,506390249;EUR;0;19954444,48;0
4	3179_GTA;3179_GTA;QUOTE B;20110630;1286556,056;2,416814711;EUR;0;53233541,23;0
5	3179_GTA;3179_GTA;QUOTE B;20110731;1821493,93;18,31717033;EUR;0;9944188,416;0
6	3179_GTA;3179_GTA;QUOTE B;20110831;1392649,835;1,624210621;EUR;0;85743179,93;
7	1657_GTA;1657_GTA;QUOTE C;20110630;256064,9989;1,414434911;EUR;0;18103696,18;
8	1657_GTA;1657_GTA;QUOTE C;20110731;307987,0783;1,62672306;EUR;0;18932975,5;0
9	1657_GTA;1657_GTA;QUOTE C;20110831;365231,97;3,873021804;EUR;0;9430155,277;0
10	2736_GTA;2736_GTA;QUOTE D;20110630;340449,0437;2,927332908;EUR;0;11630007,74;
11	2736_GTA;2736_GTA;QUOTE D;20110731;453833,7498;603,4783685;EUR;0;75202,98547;
12	2736_GTA;2736_GTA;QUOTE D;20110831;472256,4485;3,513302708;EUR;0;13441951,57;
13	7171_GTA;7171_GTA;QUOTE E;20110630;138384,029;3,072482279;EUR;0;4503981,357;0
14	7171_GTA;7171_GTA;QUOTE E;20110731;144038,8724;3,010922524;EUR;0;4783878,403;
15	7171_GTA;7171_GTA;QUOTE E;20110831;111207,8537;2,907048986;EUR;0;3825455,102;
16	1000_GTA;1000_GTA;QUOTE F;20110630;1485837,027;4,086778676;EUR;0;36357168,92;
17	1000_GTA;1000_GTA;QUOTE F;20110731;2004455,607;3,17575192;EUR;0;63117512,24;0
18	1000_GTA;1000_GTA;QUOTE F;20110831;1932787,803;4,464621042;EUR;0;43291195,04;
19	5777_GTA;5777_GTA;QUOTE G;20110630;601389,4109;1,482861857;EUR;0;40555997,05;
20	5777_GTA;5777_GTA;QUOTE G;20110731;753288,4016;1,701747622;EUR;0;44265576,85;
21	5777_GTA;5777_GTA;QUOTE G;20110831;469054,4007;2,535926518;EUR;0;18496371,94;
22	1236_GTA;1236_GTA;QUOTE H;20110630;42807,63813;3,192373797;EUR;0;1340934,391;
23	1236_GTA;1236_GTA;QUOTE I;20110731;46629,50872;1,678684454;EUR;0;2777741,141;0
24	1236_GTA;1236_GTA;QUOTE I;20110831;34669,4674;4,583577279;EUR;0;756384,4849;0
25	4767_GTA;4767_GTA;QUOTE J ;20110630;129794,935;0,989088795;EUR;0;13122677,72;0
26	4767_GTA;4767_GTA;QUOTE J;20110731;132267,1962;1,18265255;EUR;0;11183943,77;0
27	4767_GTA;4767_GTA;QUOTE J ;20110831;193076,6938;4,032790717;EUR;0;4787669,566;

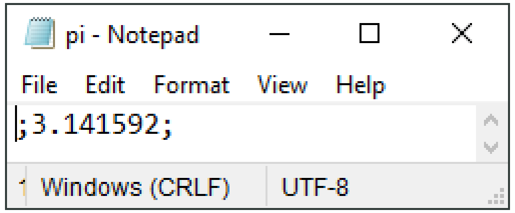
mybeautifulcsv - Notepad

File Edit Format View Help

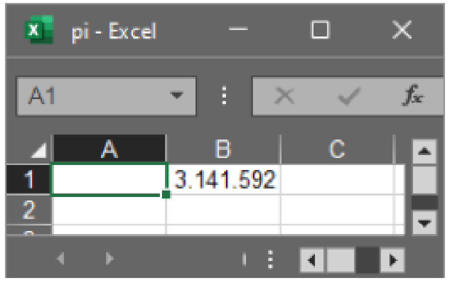
```
PK▯▯ ▯ ▯   ! bîh^▯  ▯▯  ▯ ▯[Content_Types].xml ¢▯ (

?^9LìÒ™▯ÈsbgÙ®|Èl!õù▯USh9iºbžr:"y_dlÀóD›¿▯ý|-NœÈR"4▯ø2ÏGÇ% õ▯Z´4ñËyÄ7      Ã«ÈðÉ,‹¨Þ▯  ÿÿ▯ PK▯▯ ▯ ▯   ! :▯(Õu▯  ½▯  ▯   xl/workbook.xml¬Umo£8▯þ~Òý▯ÄwŠMxWé
 ÜUx®ª4ÛÞI•V.8Å*`Î˜&UµÿýÆ$¤ífuÊu/Jlì▯?3óxrúiÓÔÚ▯▯=ãm¤ã▯¤k´-xÉÚ‡Hÿ²Ì
_xzIÚ’Ô¼¥‘þL{ýÓÙ¯¿œ®¹x¼çüQ▯€¶õJÊ.4Í¾¨hCú▯ÞÑ▯,+.▯"a)Ì¾▯”’}E©1jÓBÈ5▯ÂZ}‹▯Šc0øjÅ
šòbhh+· ,ÖD ý¾b]?¡5Å1p
▯▯Cg▯¼é âžÕL> ºÖ▯áùCË▯¹¯!ì
v´€¯▯?Œ`°¦“ÀtpTÃ
Á{¾’' mnIÄ‘‰ñ»▯l▯sp’m
úÄT
÷¬„ûAVîË}▯Ãè§Ñ0HkÔJ▯Éû š³çfég§+VÓ›-t5ÒuŸI£*UëZMz™•LÒ2Ò=Xò5}·!†.▯X
VËõfX7Ïör¾▯ZIWd¨å▯„<ÁÃÍpÝÀr”'▯#®%▯-‘tÎ[    :ÜÅõ³š▯±ç▯▯…k▯ú÷À▯…‹▯ú,Xa$EHîû+"+m▯u¤ÿ▯Þ
=„▯÷▯#mKr^x¤ä,Þ½Q&9¼▯ÿA›¤P▯›▯ñ–Õöùûèœ▯'ý]I¡Áóyz▯5¸&OP▯¨{¹»°çrÿë‹{(›§–á;ÞÌ°±ç▯qŠbcî!ß™Í▯”¢ä▯D!Ü°àdÕ®Ê
3Òm(éé’l&▯FáÀÊ×ó_Ðîc¨ù»a²}S‘ª~vÃèº▯ÕƒZj›[Ö–|é▯¶ šç÷Ëõh¼e¥¬@P▯²Áe»÷;e▯▯0ÆŽ¯6A÷ŠY¤¿Ì,Ä·íyjØÖ<6l”æF€bl▯®▯ä^,ì$†▯¨Ü¿¡4vN 6ÎZ;ªýZuS▲-ZÍ*»ð,Bu†8/Ga›Ók▯©▯P·šF
y ÑèÔ!+øÁEö▯ò¼üfMyÎkÁ£ú▲å«Kª5=|†t ‡ÈG▯)’så¢™»UïátBûÊ)¿Ûò,ËôïfäÉÇÕß   ÿÿ▯ PK▯▯ ▯ ▯   ! ØáÌó;▯  ^▯  ▯   xl/worksheets/sheet1.xmlœ“[›0▯…ß+õ?X~çb $ Uš(ê¾U
–~▯–ß`Ãç²åAÖ▯ý»ÌgIZ¤E²b▯d³e▯|ÉY▯°œ-6Óuœ&Ùú]Ìk‰;<LE,4▯]²r9™Òh1▯ôKÂÉ½«‰çÛ▯P < ▯£dÈçÖ˜ýÐøŒK1"ÝØ0 ¹ðò▯+Pª¢«)FüÏ(,%
D7…÷õUm3&ú›%54ü üwsú
r×z”MÂé▯G²RÖ¯kp CŠÚa2p…Q▯Á+ÑrøÙ0cü|1+kßVt▯‡,‹ó▯▯[p~#▯&%âà¼Ñ¿/=1t8¢þ▯  ÿÿ   ÿÿ”ØËjÂP „áW)y€ê¹y#▯
¾HH…®11bÛ·ï™Í™P2‹Y▯Q‡Ÿ,üHb?▯\¯Ëø▯‡þþùýr?w¡{"¿ÆÛ\Nõø'äq:½ÿ^®ót½=ÎÝö5vC?aúV¿¯ŸÌõýsØö›çÐo¦úª¥–‹F®n[.¬ç’‘«Û–‹ë¹läê¶åÒz®▯¹ºm¹½žÛ▯¹ºm¹²žÛ▯¹ºm¹Ýzî`äê¶åöë¹£
ÿ1ú▯  ÿÿ   ÿÿ²)HLOõM,JÏÌ+VÈIM+±U2Ð3WR(ÊLÏ€±Kò▯À¢¦J
Iù%%ù¹0^FjbJj▯ˆg¬¤–Ÿ_ ãèÛÙè–ç▯e▯g¤¦-Ø▯   ÿÿ▯ PK▯▯ ▯ ▯   ! Á▯▯¾N▯  Æ   ▯   xl/theme/theme1.xmlìYÍ‹▯7▯¿▯ú?▲swü5ã%ÞàÏl“Ý$d”µ¶ìQV32’¼▯▯ %9õR(¤¥–Bo=”Ò@▯
%ô      $´éÑ'ÍØ#-å$›lJZv
‹Gþ½§§÷ž~zótñÒ½˜zG˜▯Â’–_¾Pò=œŒØ˜$Ó–▯k8(4|OH”Œ▯e   nù▯,üKÛŸ~r▯mÉ▯ÇØ▯ùDl¡–I9Û*▯Å▯†‘¸Àf8ß&ŒÇHÂ#Ÿ▯ÇƒÞ˜▯+¥R-▯#’ø^,bP{}2!#ì
•J▯{©¼Oá1‘B
Œ(ßWª±%¡±ãÃ²B^…èRî!Úòaž1;â{Ò÷(▯▯~hù%ýç▯·/▯ÑV&Då▯YCn ÿ2¹L`|XÑsòéÁjÒ ▯ƒZ{¥_▯¨\ÇõëýZ¿¶Ò§▯h4,•¦¶Ø:ë•na
PúÕ¡»WïUË▯ÞÐ_]³¹ª…x T▯°†▲ºàE▯¯A)>\Ã‡f§gëx ▯_[Ã×Kí^P·ôkPDIr¸†.…µjw¹Ú▯dÂèŽ▯Þ▲ƒA½’)ÏQ
```

Ln 8, Col 25 | 100% | Macintosh (CR) | ANSI

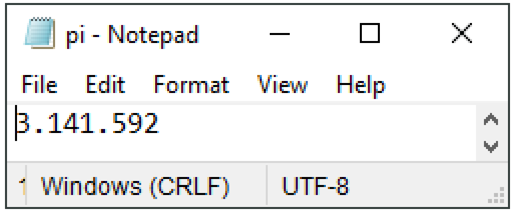
The user opens the CSV file with Excel to check or amend it, and then saves it. The data format of the fields may change (dates, decimal separator…).



```
;3.141592;
```



A	B	C
	3.141.592	



```
3.141.592
```

The import crashes or the imported data is wrong!

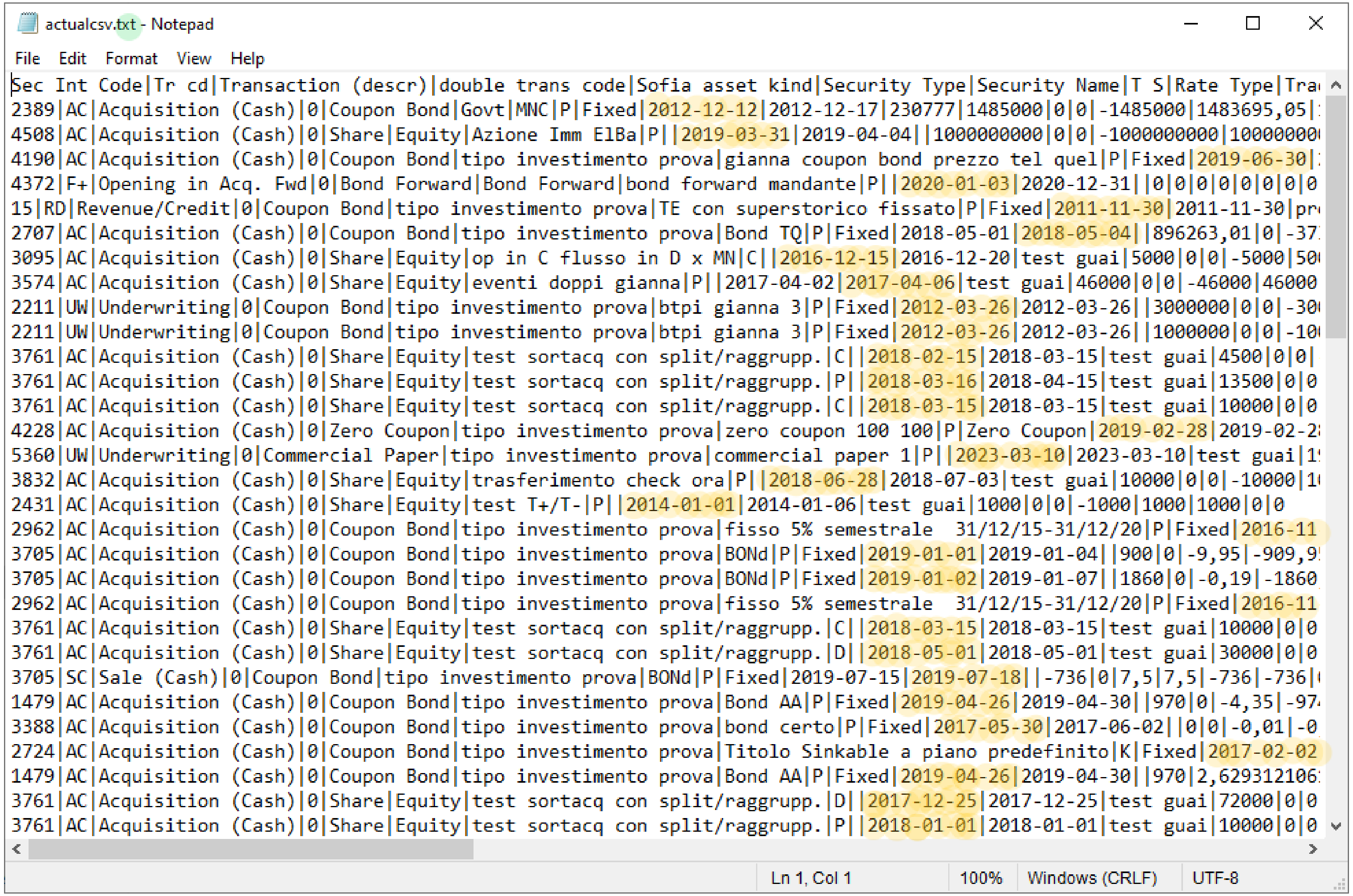
Possible solutions:

Ask the user to open the file with a text editor

Change the file extension to something else (e.g. txt), so that

Windows won’t try to open it with Excel

This is a CSV file



```
Sec Int Code|Tr cd|Transaction (descr)|double trans code|Sofia asset kind|Security Type|Security Name|T S|Rate Type|Tra
2389|AC|Acquisition (Cash)|0|Coupon Bond|Govt|MNC|P|Fixed|2012-12-12|2012-12-17|230777|1485000|0|0|-1485000|1483695,05|
4508|AC|Acquisition (Cash)|0|Share|Equity|Azione Imm ElBa|P||2019-03-31|2019-04-04||1000000000|0|0|-1000000000|10000000
4190|AC|Acquisition (Cash)|0|Coupon Bond|tipo investimento prova|gianna coupon bond prezzo tel quel|P|Fixed|2019-06-30|
4372|F+|Opening in Acq. Fwd|0|Bond Forward|Bond Forward|bond forward mandante|P||2020-01-03|2020-12-31||0|0|0|0|0|0|0|0
15|RD|Revenue/Credit|0|Coupon Bond|tipo investimento prova|TE con superstorico fissato|P|Fixed|2011-11-30|2011-11-30|pr
2707|AC|Acquisition (Cash)|0|Coupon Bond|tipo investimento prova|Bond TQ|P|Fixed|2018-05-01|2018-05-04||896263,01|0|-37
3095|AC|Acquisition (Cash)|0|Share|Equity|op in C flusso in D x MN|C||2016-12-15|2016-12-20|test guai|5000|0|0|-5000|50
3574|AC|Acquisition (Cash)|0|Share|Equity|eventi doppi gianna|P||2017-04-02|2017-04-06|test guai|46000|0|0|-46000|46000
2211|UW|Underwriting|0|Coupon Bond|tipo investimento prova|btpi gianna 3|P|Fixed|2012-03-26|2012-03-26||3000000|0|0|-30
2211|UW|Underwriting|0|Coupon Bond|tipo investimento prova|btpi gianna 3|P|Fixed|2012-03-26|2012-03-26||1000000|0|0|-10
3761|AC|Acquisition (Cash)|0|Share|Equity|test sortacq con split/raggrupp.|C||2018-02-15|2018-03-15|test guai|4500|0|0|
3761|AC|Acquisition (Cash)|0|Share|Equity|test sortacq con split/raggrupp.|P||2018-03-16|2018-04-15|test guai|13500|0|0
3761|AC|Acquisition (Cash)|0|Share|Equity|test sortacq con split/raggrupp.|C||2018-03-15|2018-03-15|test guai|10000|0|0
4228|AC|Acquisition (Cash)|0|Zero Coupon|tipo investimento prova|zero coupon 100 100|P|Zero Coupon|2019-02-28|2019-02-2
5360|UW|Underwriting|0|Commercial Paper|tipo investimento prova|commercial paper 1|P||2023-03-10|2023-03-10|test guai|1
3832|AC|Acquisition (Cash)|0|Share|Equity|trasferimento check ora|P||2018-06-28|2018-07-03|test guai|10000|0|0|-10000|1
2431|AC|Acquisition (Cash)|0|Share|Equity|test T+/T-|P||2014-01-01|2014-01-06|test guai|1000|0|0|-1000|1000|1000|0|0
2962|AC|Acquisition (Cash)|0|Coupon Bond|tipo investimento prova|fisso 5% semestrale  31/12/15-31/12/20|P|Fixed|2016-11
3705|AC|Acquisition (Cash)|0|Coupon Bond|tipo investimento prova|BONd|P|Fixed|2019-01-01|2019-01-04||900|0|-9,95|-909,9
3705|AC|Acquisition (Cash)|0|Coupon Bond|tipo investimento prova|BONd|P|Fixed|2019-01-02|2019-01-07||1860|0|-0,19|-1860
2962|AC|Acquisition (Cash)|0|Coupon Bond|tipo investimento prova|fisso 5% semestrale  31/12/15-31/12/20|P|Fixed|2016-11
3761|AC|Acquisition (Cash)|0|Share|Equity|test sortacq con split/raggrupp.|C||2018-03-15|2018-03-15|test guai|10000|0|0
3761|AC|Acquisition (Cash)|0|Share|Equity|test sortacq con split/raggrupp.|D||2018-05-01|2018-05-01|test guai|30000|0|0
3705|SC|Sale (Cash)|0|Coupon Bond|tipo investimento prova|BONd|P|Fixed|2019-07-15|2019-07-18||-736|0|7,5|7,5|-736|-736|
1479|AC|Acquisition (Cash)|0|Coupon Bond|tipo investimento prova|Bond AA|P|Fixed|2019-04-26|2019-04-30||970|0|-4,35|-97
3388|AC|Acquisition (Cash)|0|Coupon Bond|tipo investimento prova|bond certo|P|Fixed|2017-05-30|2017-06-02||0|0|-0,01|-0
2724|AC|Acquisition (Cash)|0|Coupon Bond|tipo investimento prova|Titolo Sinkable a piano predefinito|K|Fixed|2017-02-02
1479|AC|Acquisition (Cash)|0|Coupon Bond|tipo investimento prova|Bond AA|P|Fixed|2019-04-26|2019-04-30||970|2,629312106
3761|AC|Acquisition (Cash)|0|Share|Equity|test sortacq con split/raggrupp.|D||2017-12-25|2017-12-25|test guai|72000|0|0
3761|AC|Acquisition (Cash)|0|Share|Equity|test sortacq con split/raggrupp.|P||2018-01-01|2018-01-01|test guai|10000|0|0
```

At any rate, it isn’t easy to read it with a text editor!

UTF-8

1 char is not always 1 byte (multibyte encoding)

Reading files in chunks has become more difficult

Bite Order Mark (BOM)

It is more work for us, but

non-ANSI characters do exist

BOM can even help us in some cases

A tale of two encodings

The task:

Customized ANSI export for an archive, had to be updated to UTF-8 for special characters support

Sofia must be able to read both old and new files, but we didn’t want to convert all old files

The execution:

From ⎕NREAD to ⎕NGET

⎕NGET is used with the encoding for old files

New files are UTF-8 with BOM: when ⎕NGET detects BOM, it uses the BOM encoding even if the encoding is provided in the left argument

Thanks ⎕NGET + BOM!

XML

Extensively used for web application

Tags and hierarchy should improve user readibility

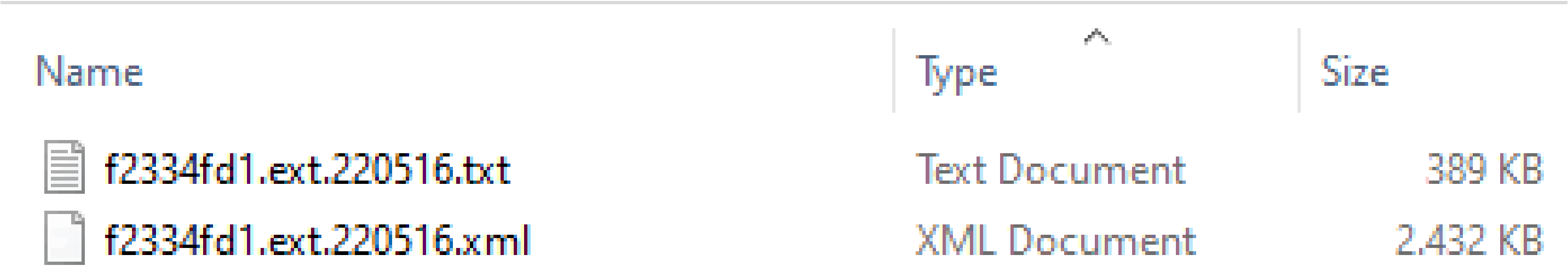
Typically, all information is stored in the file, so that in can be used by more applications for different purpuses

Really easy to manage thanks to □XML

But

Reading file in chunks is possible but requires some tricks

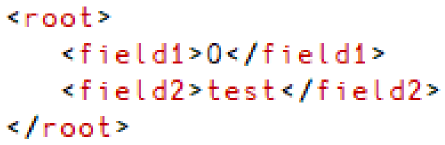
To exchange the same data you need bigger files
(also compared to fixed width files)



Name	Type	Size
f2334fd1.ext.220516.txt	Text Document	389 KB
f2334fd1.ext.220516.xml	XML Document	2.432 KB

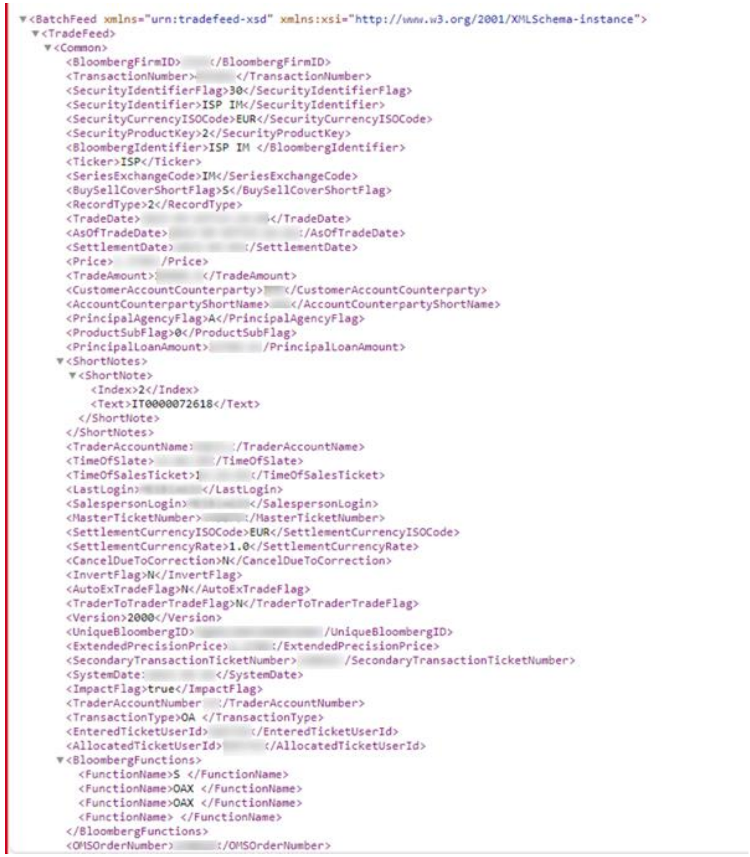
User readibility

This is neat!



```
<root>
   <field1>0</field1>
   <field2>test</field2>
</root>
```

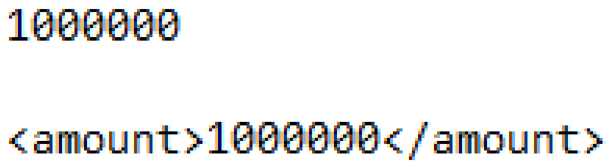
This is somewhat more challenging



```
▼<BatchFeed xmlns="urn:tradefeed-xsd" xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance">
  ▼<TradeFeed>
    ▼<Common>
      <BloombergFirmID>[redacted]</BloombergFirmID>
      <TransactionNumber>[redacted]</TransactionNumber>
      <SecurityIdentifierFlag>30</SecurityIdentifierFlag>
      <SecurityIdentifier>ISP IM</SecurityIdentifier>
      <SecurityCurrencyISOCode>EUR</SecurityCurrencyISOCode>
      <SecurityProductKey>2</SecurityProductKey>
      <BloombergIdentifier>ISP IM </BloombergIdentifier>
      <Ticker>ISP</Ticker>
      <SeriesExchangeCode>IM</SeriesExchangeCode>
      <BuySellCoverShortFlag>S</BuySellCoverShortFlag>
      <RecordType>2</RecordType>
      <TradeDate>[redacted]</TradeDate>
      <AsOfTradeDate>[redacted]/AsOfTradeDate>
      <SettlementDate>[redacted]/SettlementDate>
      <Price>[redacted]/Price>
      <TradeAmount>[redacted]</TradeAmount>
      <CustomerAccountCounterparty>[redacted]</CustomerAccountCounterparty>
      <AccountCounterpartyShortName>[redacted]</AccountCounterpartyShortName>
      <PrincipalAgencyFlag>A</PrincipalAgencyFlag>
      <ProductSubFlag>0</ProductSubFlag>
      <PrincipalLoanAmount>[redacted]/PrincipalLoanAmount>
    ▼<ShortNotes>
      ▼<ShortNote>
          <Index>2</Index>
          <Text>IT0000072618</Text>
        </ShortNote>
      </ShortNotes>
      <TraderAccountName>[redacted]/TraderAccountName>
      <TimeOfSlate>[redacted]/TimeOfSlate>
      <TimeOfSalesTicket>[redacted]</TimeOfSalesTicket>
      <LastLogin>[redacted]</LastLogin>
      <SalespersonLogin>[redacted]</SalespersonLogin>
      <MasterTicketNumber>[redacted]/MasterTicketNumber>
      <SettlementCurrencyISOCode>EUR</SettlementCurrencyISOCode>
      <SettlementCurrencyRate>1.0</SettlementCurrencyRate>
      <CancelDueToCorrection>N</CancelDueToCorrection>
      <InvertFlag>N</InvertFlag>
      <AutoExTradeFlag>N</AutoExTradeFlag>
      <TraderToTraderTradeFlag>N</TraderToTraderTradeFlag>
      <Version>2000</Version>
      <UniqueBloombergID>[redacted]/UniqueBloombergID>
      <ExtendedPrecisionPrice>[redacted]/ExtendedPrecisionPrice>
      <SecondaryTransactionTicketNumber>[redacted]/SecondaryTransactionTicketNumber>
      <SystemDate[redacted]</SystemDate>
      <ImpactFlag>true</ImpactFlag>
      <TraderAccountNumber[redacted]/TraderAccountNumber>
      <TransactionType>OA </TransactionType>
      <EnteredTicketUserId>[redacted]</EnteredTicketUserId>
      <AllocatedTicketUserId>[redacted]</AllocatedTicketUserId>
    ▼<BloombergFunctions>
        <FunctionName>S </FunctionName>
        <FunctionName>OAX </FunctionName>
        <FunctionName>OAX </FunctionName>
        <FunctionName> </FunctionName>
      </BloombergFunctions>
      <OMSOrderNumber>[redacted]/OMSOrderNumber>
```

Why does this happen?

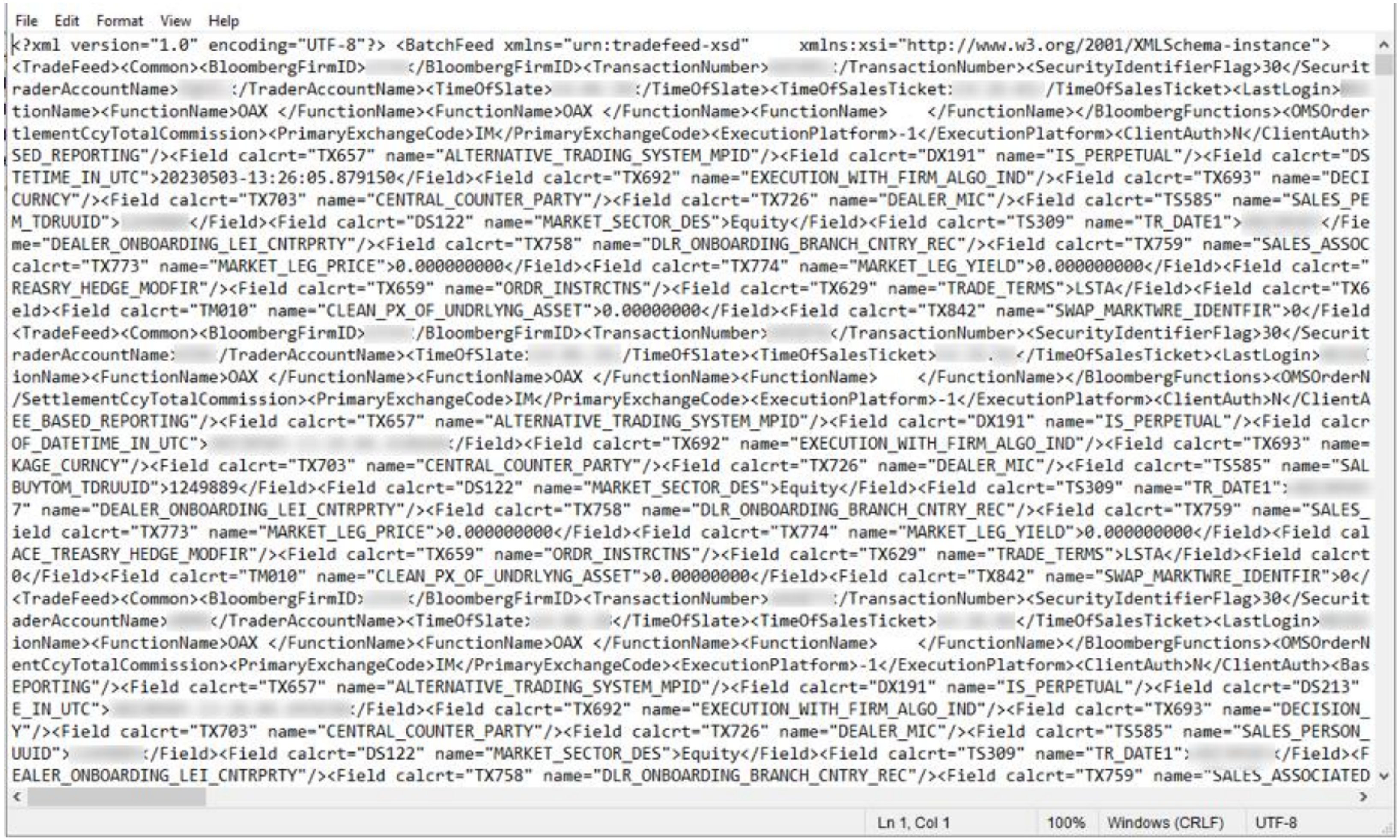
By design: XML requires to open and close the tag, while for csv you simply need the delimiter



```
1000000

<amount>1000000</amount>
```

Not tailored export: all available information is included in the file, so that it can satisfy all possible present and future needs

When the extension of the XML file is txt, the file is automatically opened with Notepad



```
<?xml version="1.0" encoding="UTF-8"?> <BatchFeed xmlns="urn:tradefeed-xsd"     xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance">
<TradeFeed><Common><BloombergFirmID>[redacted]</BloombergFirmID><TransactionNumber>[redacted]</TransactionNumber><SecurityIdentifierFlag>30</Securit
raderAccountName>[redacted]</TraderAccountName><TimeOfSlate>[redacted]</TimeOfSlate><TimeOfSalesTicket>[redacted]/TimeOfSalesTicket><LastLogin>[redacted]
tionName><FunctionName>OAX </FunctionName><FunctionName>OAX </FunctionName><FunctionName>    </FunctionName></BloombergFunctions><OMSOrder
tlementCcyTotalCommission><PrimaryExchangeCode>IM</PrimaryExchangeCode><ExecutionPlatform>-1</ExecutionPlatform><ClientAuth>N</ClientAuth>
SED_REPORTING"/><Field calcrt="TX657" name="ALTERNATIVE_TRADING_SYSTEM_MPID"/><Field calcrt="DX191" name="IS_PERPETUAL"/><Field calcrt="DS
TETIME_IN_UTC">20230503-13:26:05.879150</Field><Field calcrt="TX692" name="EXECUTION_WITH_FIRM_ALGO_IND"/><Field calcrt="TX693" name="DECI
CURNCY"/><Field calcrt="TX703" name="CENTRAL_COUNTER_PARTY"/><Field calcrt="TX726" name="DEALER_MIC"/><Field calcrt="TS585" name="SALES_PE
M_TDRUUID">[redacted]</Field><Field calcrt="DS122" name="MARKET_SECTOR_DES">Equity</Field><Field calcrt="TS309" name="TR_DATE1">[redacted]</Fie
me="DEALER_ONBOARDING_LEI_CNTRPRTY"/><Field calcrt="TX758" name="DLR_ONBOARDING_BRANCH_CNTRY_REC"/><Field calcrt="TX759" name="SALES_ASSOC
calcrt="TX773" name="MARKET_LEG_PRICE">0.000000000</Field><Field calcrt="TX774" name="MARKET_LEG_YIELD">0.000000000</Field><Field calcrt="
REASRY_HEDGE_MODFIR"/><Field calcrt="TX659" name="ORDR_INSTRCTNS"/><Field calcrt="TX629" name="TRADE_TERMS">LSTA</Field><Field calcrt="TX6
eld><Field calcrt="TM010" name="CLEAN_PX_OF_UNDRLYNG_ASSET">0.00000000</Field><Field calcrt="TX842" name="SWAP_MARKTWRE_IDENTFIR">0</Field
<TradeFeed><Common><BloombergFirmID>[redacted]/BloombergFirmID><TransactionNumber>[redacted]</TransactionNumber><SecurityIdentifierFlag>30</Securit
raderAccountName>[redacted]/TraderAccountName><TimeOfSlate>[redacted]/TimeOfSlate><TimeOfSalesTicket>[redacted]</TimeOfSalesTicket><LastLogin>[redacted]
ionName><FunctionName>OAX </FunctionName><FunctionName>OAX </FunctionName><FunctionName>    </FunctionName></BloombergFunctions><OMSOrderN
/SettlementCcyTotalCommission><PrimaryExchangeCode>IM</PrimaryExchangeCode><ExecutionPlatform>-1</ExecutionPlatform><ClientAuth>N</ClientA
EE_BASED_REPORTING"/><Field calcrt="TX657" name="ALTERNATIVE_TRADING_SYSTEM_MPID"/><Field calcrt="DX191" name="IS_PERPETUAL"/><Field calcr
OF_DATETIME_IN_UTC">[redacted]</Field><Field calcrt="TX692" name="EXECUTION_WITH_FIRM_ALGO_IND"/><Field calcrt="TX693" name=
KAGE_CURNCY"/><Field calcrt="TX703" name="CENTRAL_COUNTER_PARTY"/><Field calcrt="TX726" name="DEALER_MIC"/><Field calcrt="TS585" name="SAL
BUYTOM_TDRUUID">1249889</Field><Field calcrt="DS122" name="MARKET_SECTOR_DES">Equity</Field><Field calcrt="TS309" name="TR_DATE1">[redacted]
7" name="DEALER_ONBOARDING_LEI_CNTRPRTY"/><Field calcrt="TX758" name="DLR_ONBOARDING_BRANCH_CNTRY_REC"/><Field calcrt="TX759" name="SALES_
ield calcrt="TX773" name="MARKET_LEG_PRICE">0.000000000</Field><Field calcrt="TX774" name="MARKET_LEG_YIELD">0.000000000</Field><Field cal
ACE_TREASRY_HEDGE_MODFIR"/><Field calcrt="TX659" name="ORDR_INSTRCTNS"/><Field calcrt="TX629" name="TRADE_TERMS">LSTA</Field><Field calcrt
0</Field><Field calcrt="TM010" name="CLEAN_PX_OF_UNDRLYNG_ASSET">0.00000000</Field><Field calcrt="TX842" name="SWAP_MARKTWRE_IDENTFIR">0</
<TradeFeed><Common><BloombergFirmID>[redacted]</BloombergFirmID><TransactionNumber>[redacted]</TransactionNumber><SecurityIdentifierFlag>30</Securit
aderAccountName>[redacted]</TraderAccountName><TimeOfSlate>[redacted]</TimeOfSlate><TimeOfSalesTicket>[redacted]</TimeOfSalesTicket><LastLogin>[redacted]
ionName><FunctionName>OAX </FunctionName><FunctionName>OAX </FunctionName><FunctionName>    </FunctionName></BloombergFunctions><OMSOrderN
entCcyTotalCommission><PrimaryExchangeCode>IM</PrimaryExchangeCode><ExecutionPlatform>-1</ExecutionPlatform><ClientAuth>N</ClientAuth><Bas
EPORTING"/><Field calcrt="TX657" name="ALTERNATIVE_TRADING_SYSTEM_MPID"/><Field calcrt="DX191" name="IS_PERPETUAL"/><Field calcrt="DS213"
E_IN_UTC">[redacted]</Field><Field calcrt="TX692" name="EXECUTION_WITH_FIRM_ALGO_IND"/><Field calcrt="TX693" name="DECISION_
Y"/><Field calcrt="TX703" name="CENTRAL_COUNTER_PARTY"/><Field calcrt="TX726" name="DEALER_MIC"/><Field calcrt="TS585" name="SALES_PERSON_
UUID">[redacted]</Field><Field calcrt="DS122" name="MARKET_SECTOR_DES">Equity</Field><Field calcrt="TS309" name="TR_DATE1">[redacted]</Field><F
EALER_ONBOARDING_LEI_CNTRPRTY"/><Field calcrt="TX758" name="DLR_ONBOARDING_BRANCH_CNTRY_REC"/><Field calcrt="TX759" name="SALES_ASSOCIATED
```

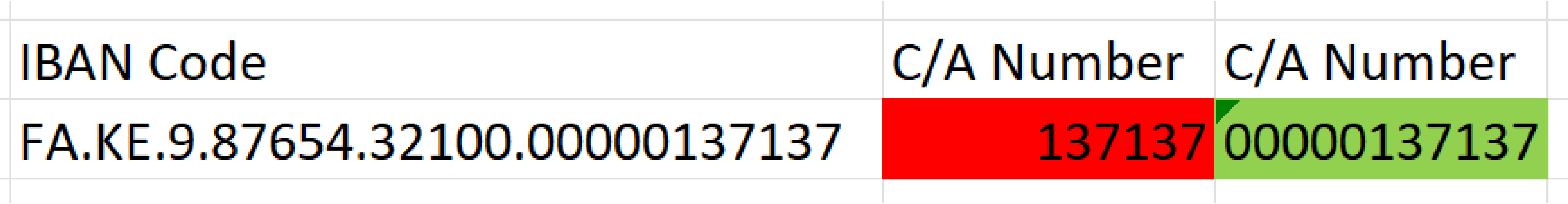
Excel

The preferred solution when users directly create the file: human to software interaction

Every field can have a specific format (dates, strings, numbers)

Field with general format are converted (if possible) when opening/reading the file

- ⚠ Beware strings of numbers (e.g. C/A numbers are treated as strings, and they often begin with a series of 0)

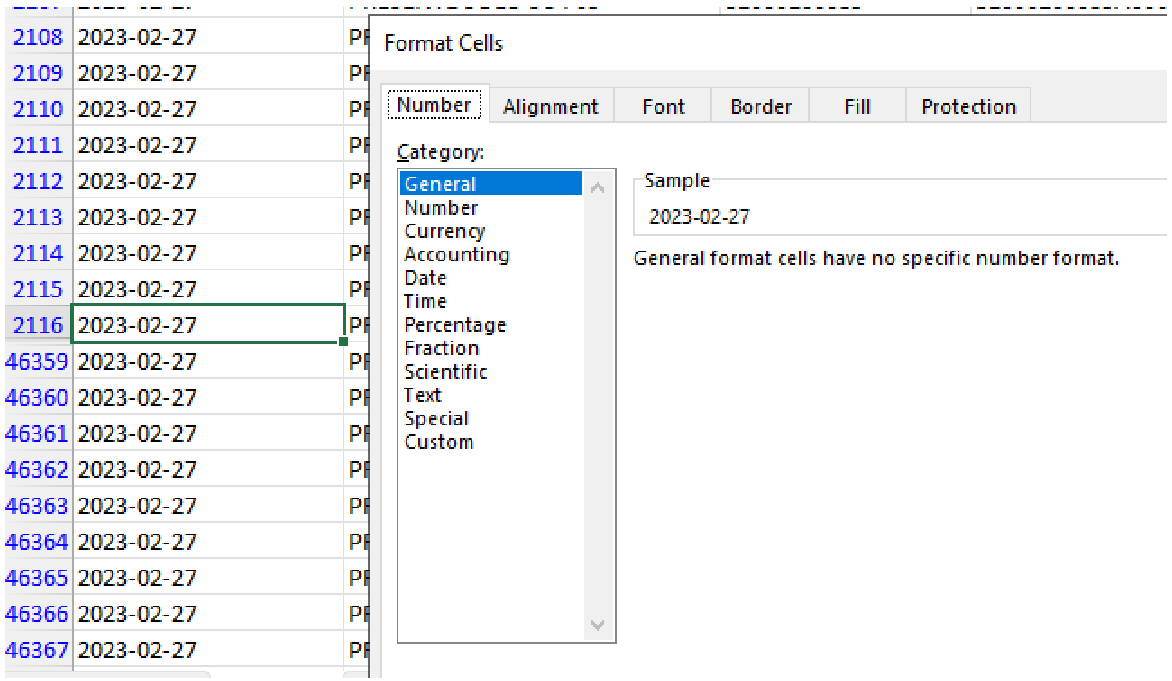


IBAN Code	C/A Number	C/A Number
FA.KE.9.87654.32100.00000137137	137137	00000137137

All that glitters is not gold

The content of Excel fields depend on regional settings of the Windows session it has been run on

E.g. if date fields are not in date format, they are convert to dates only if the format is the same of local settings



2108	2023-02-27	PF
2109	2023-02-27	PF
2110	2023-02-27	PF
2111	2023-02-27	PF
2112	2023-02-27	PF
2113	2023-02-27	PF
2114	2023-02-27	PF
2115	2023-02-27	PF
2116	2023-02-27	PF
46359	2023-02-27	PF
46360	2023-02-27	PF
46361	2023-02-27	PF
46362	2023-02-27	PF
46363	2023-02-27	PF
46364	2023-02-27	PF
46365	2023-02-27	PF
46366	2023-02-27	PF
46367	2023-02-27	PF

Format Cells

Category: General

Sample

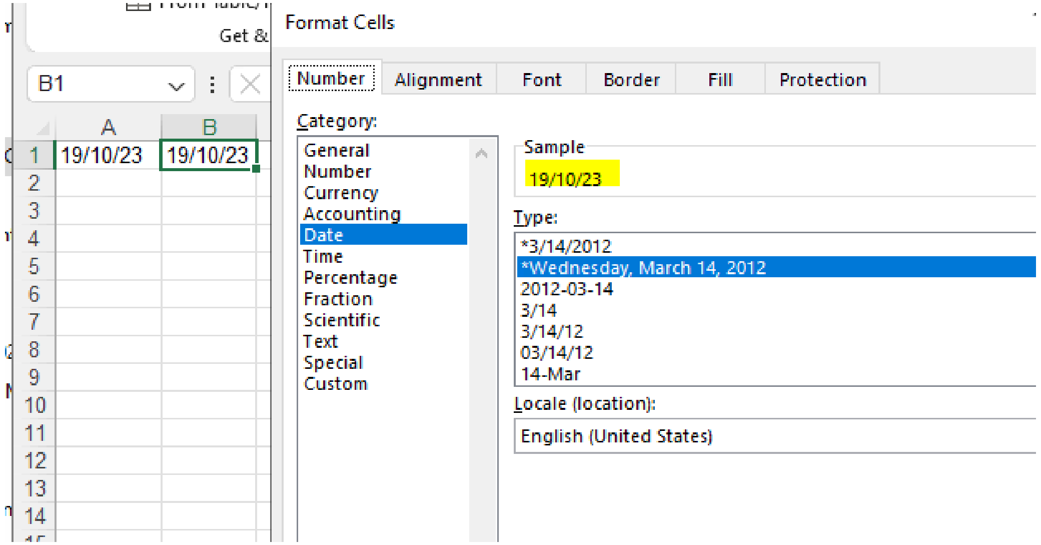
2023-02-27

General format cells have no specific number format.

A little digression

Dates are always something to manage carefully

19/10/23 that on an Italian server is read as October 19th 2023 but on an English (United States) server is only a string



	A	B
1	19/10/23	19/10/23

Category:

General
Number
Currency
Accounting
Date
Time
Percentage
Fraction
Scientific
Text
Special
Custom

Sample

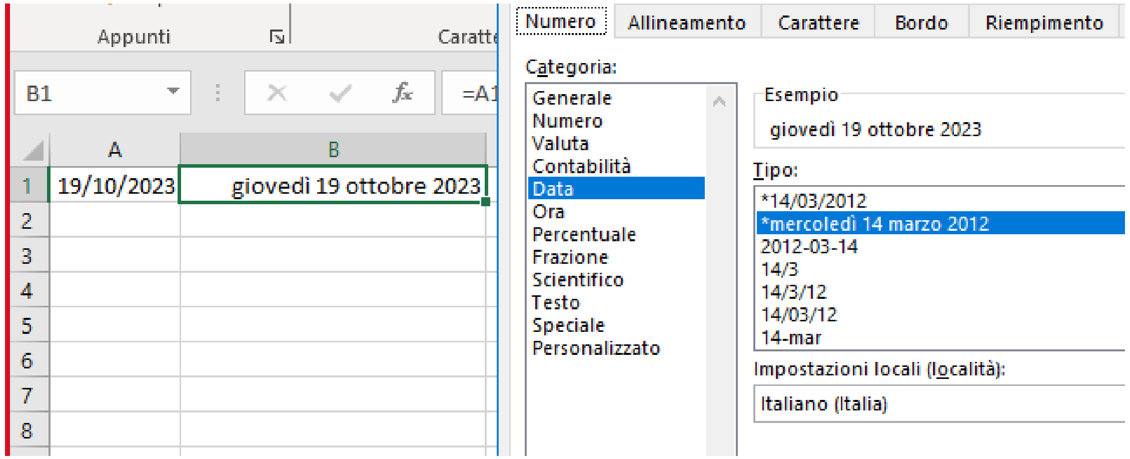
19/10/23

Type:

*3/14/2012
*Wednesday, March 14, 2012
2012-03-14
3/14
3/14/12
03/14/12
14-Mar

Locale (location):

English (United States)



	A	B
1	19/10/2023	giovedì 19 ottobre 2023

Categoria:

Generale
Numero
Valuta
Contabilità
Data
Ora
Percentuale
Frazione
Scientifico
Testo
Speciale
Personalizzato

Esempio

giovedì 19 ottobre 2023

Tipo:

*14/03/2012
*mercoledì 14 marzo 2012
2012-03-14
14/3
14/3/12
14/03/12
14-mar

Impostazioni locali (località):

Italiano (Italia)

5/4/23

April 5th,2023

May the force!!!

Using Excel to read and write Excel files

Imports and exports crash if Excel has problems, e.g. already in use on the server or not properly installed/authorized

Solution: close all pending Excel session and/or fix installation/authorization (unfortunately there isn’t something else to do)

Different versions of Excel can cause exports to be blocked

Solution: create new file or save a copy with the newer version

In the last years we are using a DLL to read Excel files, so there is no longer the need to have Excel installed for imports

But another issue popped up:

- the things we want to read are formulas that need to be evaluated
- the values we want to import don’t really exist until Excel evaluate them
- we cannot import the numbers we need

Solution: fortunately, there was a method in DLL to evaluate formulas before reading values.

So, how to deal when things with files go wrong?

Curiosity

Collaboration

Patience

Patience