



worst, but still usable, data I've confronted

**Dr. Arthur Tabachneck
Director, Data Management**



suppose you had the following excel file:

	A	B	C	D	E	F	G
1	date						
2	jan.1, 2009	← formatted as text					
3	Friday, January 02, 2009	← formatted as shown					
4	Saturday, January 03, 2009						
5	01/04/09						
6	01/05/09	← formatted as m/d/yyyy					
7	01/06/09						
8	7/1/2009						
9	8/1/2009	← formatted as d/m/yyyy					
10	9/1/2009						
11	jan-10-09	← formatted as mon-d-yy					
12	11-Jan	← formatted as d-mon					
13	january-12-2009	← formatted as text					
14	january/13/2009						
15	2009-jan-14	← formatted as yyyy-mon-d					



how the file got so bad:

- members of a secretarial pool were asked to enter the data while they were covering the front desk**
- they (four different secretaries), obviously, weren't given sufficient instructions**
- their task was simply to enter some data, including a date**



proc import can't do it directly

it will return missing values for the following rows:

	A	B	C	D	E	F	G
1	date						
2	jan.1, 2009						
3	Friday, January 02, 2009						
4	Saturday, January 03, 2009						
5	01/04/09						
6	01/05/09						
7	01/06/09						
8	7/1/2009						
9	8/1/2009						
10	9/1/2009						
11	jan-10-09						
12		11-Jan					
13	january-12-2009						
14	january/13/2009						
15	2009-jan-14						



but it can do it in three steps

step 1: use mixed=no

```
Log: (Untitled)
NOTE: Copyright (c) 2002-2003 by SAS Institute Inc., Cary, NC, USA.
NOTE: SAS (r) 9.1 (TS1M3)
NOTE: Licensed to INSURANCE BUREAU OF CANADA, Site 0022719002.
NOTE: This session is executing on the NET_ASRV platform.

NOTE: SAS 9.1.3 Service Pack 9
NOTE: SAS initialization used:

testdateppt.sas
/*Import the records that SAS can read as a date -- others will be missing*/
PROC IMPORT OUT= WORK.INPUTa
  DATAFILE= "c:\DateTest.xls"
  DBMS=EXCEL REPLACE;
  SHEET="Sheet1$";
  GETNAMES=YES;
  MIXED=NO;
  SCANTEXT=YES;
  USEDATE=YES;
  SCANTIME=YES;
RUN;
```



but it can do it in three steps

step 1: imports date formatted cells and assigns missing to other cells

	A	B	C	D	E	F	G
1	date						
2	jan.1, 2009						
3	Friday, January 02, 2009	←					
4	Saturday, January 03, 2009	←					
5	01/04/09	←					
6	01/05/09	←					
7	01/06/09	←					
8	7/1/2009	←					
9	8/1/2009	←					
10	9/1/2009	←					
11	jan-10-09						
12		11-Jan ←					
13	january-12-2009						
14	january/13/2009						
15	2009-jan-14						



but it can do it in three steps

step 2: use mixed=yes

```
Log - (Untitled)
NOTE: Copyright (c) 2002-2003 by SAS Institute Inc., Cary, NC, USA.
NOTE: SAS (r) 9.1 (TS1M3)
NOTE: Licensed to INSURANCE BUREAU OF CANADA, Site 0022719002.
NOTE: This session is executing on the NET_ASRV platform.

NOTE: SAS 9.1.3 Service Pack 3
NOTE: SAS initialization used:

testdateppt.sas
/*Import the other records*/
PROC IMPORT OUT= WORK.INPUTb
  DATAFILE= "c:\DateTest.xls"
  DBMS=EXCEL REPLACE;
  SHEET="Sheet1$";
  GETNAMES=YES;
  MIXED=YES;
  SCANTEXT=YES;
  USEDATE=YES;
  SCANTIME=YES;
RUN;
```



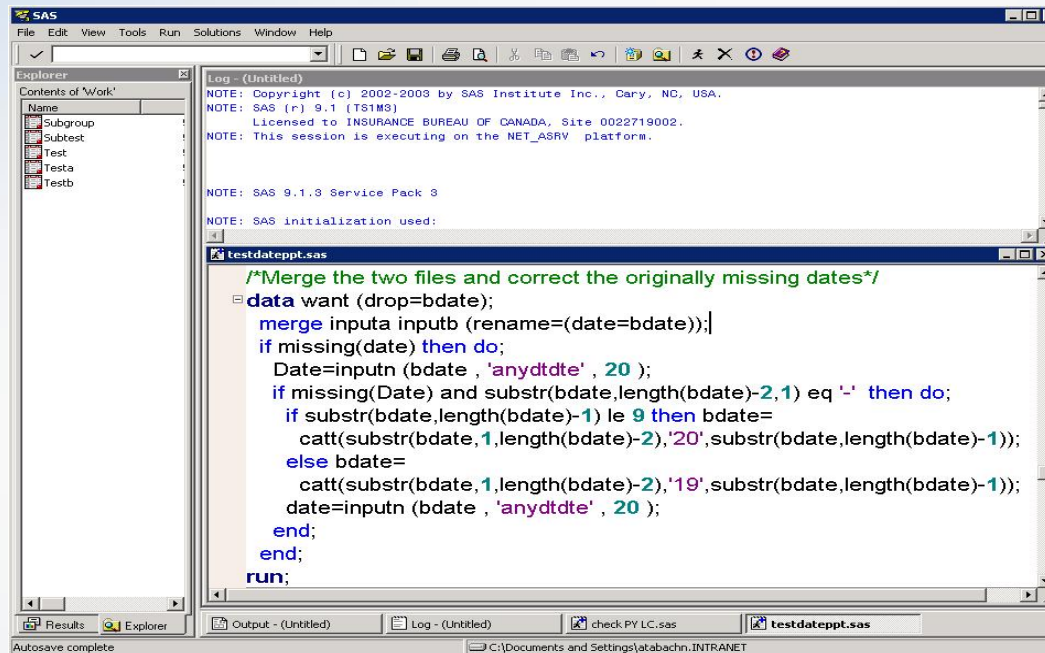
but it can do it in three steps

step 2: which will import all cells as text

	A	B	C	D	E	F	G
1	date						
2	jan.1, 2009						
3	Friday, January 02, 2009						
4	Saturday, January 03, 2009						
5	01/04/09						
6	01/05/09						
7	01/06/09						
8	7/1/2009						
9	8/1/2009						
10	9/1/2009						
11	jan-10-09						
12	11-Jan						
13	january-12-2009						
14	january/13/2009						
15	2009-jan-14						

but it can do it in three steps

step 3: use data step merge



The screenshot shows the SAS software interface. The main window displays a SAS program named 'testdatept.sas'. The program code is as follows:

```
/*Merge the two files and correct the originally missing dates*/  
data want (drop=bdate);  
  merge inputa inputb (rename=(date=bdate));  
  if missing(date) then do;  
    Date=inputn (bdate , 'anydtdte' , 20 );  
    if missing(Date) and substr(bdate,length(bdate)-2,1) eq '-' then do;  
      if substr(bdate,length(bdate)-1) le 9 then bdate=  
        catt(substr(bdate,1,length(bdate)-2),'20',substr(bdate,length(bdate)-1));  
    else bdate=  
      catt(substr(bdate,1,length(bdate)-2),'19',substr(bdate,length(bdate)-1));  
    date=inputn (bdate , 'anydtdte' , 20 );  
  end;  
end;  
run;
```

The interface also shows a log window with the following text:

```
Log - (Untitled)  
NOTE: Copyright (c) 2002-2003 by SAS Institute Inc., Cary, NC, USA.  
NOTE: SAS (r) 9.1 (TS1M3)  
       Licensed to INSURANCE BUREAU OF CANADA, Site 0022719002.  
NOTE: This session is executing on the NET_ASRV platform.  
  
NOTE: SAS 9.1.3 Service Pack 3  
  
NOTE: SAS initialization used:
```



resulting in the following file

	date
1	01JAN2009
2	02JAN2009
3	03JAN2009
4	04JAN2009
5	05JAN2009
6	06JAN2009
7	07JAN2009
8	08JAN2009
9	09JAN2009
10	10JAN2009
11	11JAN2009
12	12JAN2009
13	13JAN2009
14	14JAN2009



Questions?



But what if I don't license SAS/Access for PC Files?



you can do it with DDE

The screenshot shows the SAS software interface. The top window is the 'Log - (Untitled)' window, which contains the following text:

```
NOTE: Copyright (c) 2002-2003 by SAS Institute Inc., Cary, NC, USA.  
NOTE: SAS (r) 9.1 (TS1M3)  
       Licensed to INSURANCE BUREAU OF CANADA, Site 0022719002.  
NOTE: This session is executing on the NET_ASRV platform.  
  
NOTE: SAS 9.1.3 Service Pack 3
```

The bottom window is the 'testdateppt.sas *' code editor, which contains the following SAS code:

```
/*Set options and filename for dde commands*/  
  
options noxsync noxwait xmin;  
filename sas2xl dde 'excel|system';
```

The status bar at the bottom of the window displays 'File saved successfully.' and the path 'C:\Documents and Settings\atabachn.INTRANET'.



a dde import that does the job quite nicely

The screenshot shows the SAS software interface. The top window is the 'Log - (Untitled)' window, which contains the following text:

```
NOTE: Copyright (c) 2002-2003 by SAS Institute Inc., Cary, NC, USA.  
NOTE: SAS (r) 9.1 (TS1M3)  
       Licensed to INSURANCE BUREAU OF CANADA, Site 0022719002.  
NOTE: This session is executing on the NET_ASRV platform.  
  
NOTE: SAS 9.1.3 Service Pack 3
```

The bottom window is the 'testdateppt.sas *' code editor, which contains the following SAS code:

```
/*Open Excel*/  
data _null_;  
  length fid rc start stop time 8;  
  fid=fopen('sas2xl','s');  
  if (fid le 0) then do;  
    rc=system('start excel');  
    start=datetime();  
    stop=start+10;  
    do while (fid le 0);  
      fid=fopen('sas2xl','s');  
      time=datetime();  
      if (time ge stop) then fid=1;  
    end;  
  end;  
  rc=fclose(fid);  
run;
```

The SAS interface also shows a menu bar (File, Edit, View, Tools, Run, Solutions, Window, Help), a toolbar, and a taskbar at the bottom with the following tabs: Results, Explorer, Output - (Untitled), Log - (Untitled), and testdateppt.sas *. The status bar at the bottom indicates 'Autosave complete' and the path 'C:\Documents and Settings\atabachn.INTRANET'.



a dde import that does the job quite nicely

The screenshot shows the SAS software interface. The top menu bar includes File, Edit, View, Tools, Run, Solutions, Window, and Help. Below the menu bar is a toolbar with various icons. On the left, there is an Explorer window showing the contents of the 'work' directory. The main window is divided into two panes. The top pane is a log window titled 'Log - (Untitled)' containing the following text:

```
NOTE: Copyright (c) 2002-2003 by SAS Institute Inc., Cary, NC, USA.  
NOTE: SAS (r) 9.1 (TS1M3)  
       Licensed to INSURANCE BUREAU OF CANADA, Site 0022719002.  
NOTE: This session is executing on the NET_ASRV platform.  
  
NOTE: SAS 9.1.3 Service Pack 3
```

The bottom pane is a code editor window titled 'testdatepnt.sas *' containing the following SAS code:

```
/*Open spreadsheet*/  
  
data _null_;  
  file sas2xl;  
  put 'open("c:\datetest.xls")';  
run;  
  
/*Insert an old-style macro-sheet into the workbook.*/  
  
data _null_;  
  file sas2xl;  
  put '[workbook.next()];'  
  put '[workbook.insert(3)]';  
run;
```

The bottom status bar shows 'Autosave complete' and the path 'C:\Documents and Settings\atabachn\INTRANET'.



a dde import that does the job quite nicely

```

SAS
File Edit View Tools Run Solutions Window Help
Log - (Untitled)
NOTE: Copyright (c) 2002-2003 by SAS Institute Inc., Cary, NC, USA.
NOTE: SAS (r) 9.1 (TS1M3)
      Licensed to INSURANCE BUREAU OF CANADA, Site 0022719002.
NOTE: This session is executing on the NET_ASRV platform.

NOTE: SAS 9.1.3 Service Pack 3

testdateppt.sas *
/*Create and run the macro*/
filename xlmacro dde 'excel|macro1!r1c1:r100c1' notab lrecl=200;
data _null_;
  file xlmacro;
  put '=set.name("Tag",!$b$1)';
  put '=formula("<>",Tag)';
  put '=set.name("OldValue",!$c$1)';
  put '=set.name("NewValue",!$b$2)';
  put '=for.cell("CurrentCell",sheet1!$a$2:$a$100,true)';
  put '=formula(get.cell(5,CurrentCell),OldValue)';
  put '=formula("=concatenate(Tag,OldValue)",NewValue)';
  put '=formula(NewValue,CurrentCell)';
  put '=next()';
  put '=halt(true)';
  put '!dde_flush';
  file sas2xl;
  put "[run("macro1!r1c1")]";
run;
filename xlmacro clear;
  
```




a dde import that does the job quite nicely

```
NOTE: Copyright (c) 2002-2003 by SAS Institute Inc., Cary, NC, USA.
NOTE: SAS (r) 9.1 (TS1M3)
      Licensed to INSURANCE BUREAU OF CANADA, Site 0022719002.
NOTE: This session is executing on the NET_ASRV platform.

NOTE: SAS 9.1.3 Service Pack 3
NOTE: SAS initialization used:

/*Save the spreadsheet as a csv file - then import the data*/
data want (keep=date);
  infile "c:\DateTest.csv" dsd dlm="," lrecl=32768 firstobs=2;
  informat rawdate $20.;
  input rawdate;
  format date date9.;
  rawdate=substr(rawdate,3);
  if anyalpha(rawdate) then do;
    date=inputn (rawdate , 'anydtdte' , 20 );
    if missing(date) and substr(rawdate,length(rawdate)-2,1) eq '-' then do;
      if substr(rawdate,length(rawdate)-1) le 9 then rawdate=
        catt(substr(rawdate,1,length(rawdate)-2),'20',substr(rawdate,length(rawdate)-1));
    else rawdate=
      catt(substr(rawdate,1,length(rawdate)-2),'19',substr(rawdate,length(rawdate)-1));
    date=inputn (rawdate , 'anydtdte' , 20 );
  end;
end;
else Date=rawdate-21916;
run;
```



resulting in the following file

	date
1	01JAN2009
2	02JAN2009
3	03JAN2009
4	04JAN2009
5	05JAN2009
6	06JAN2009
7	07JAN2009
8	08JAN2009
9	09JAN2009
10	10JAN2009
11	11JAN2009
12	12JAN2009
13	13JAN2009
14	14JAN2009



Questions?

Your comments and questions are valued and encouraged.

Contact the author:

Dr. Arthur Tabachneck
Director, Data Management
Insurance Bureau of Canada
Toronto, Ontario L3T 5K9
Email: atabachneck@ibc.ca