

Beyond the Comma Delimited File for Bringing Data into a SAS Dataset

Paper CC13

David Franklin
Independent SAS Consultant
TheProgrammersCabin.com

Some trivia

“Q” is the only letter in the alphabet that does
not appear in the name of any state of the
United States.

Paper CC13, PharmaSUG2011 : David Franklin

Introduction

- Motivation: "Help! I got this AE data in a Word document and I need it in a SAS® dataset within the hour!"
- SAS has developed a number of tools over the years like Microsoft Excel and Access with PROC IMPORT, but there are times when this will not do the job
- It is back to the INFILE and INPUT statements within a datastep

Lets look at the basic styles of input, Part 1

- **List Input**

```
input city $ distance;  
cards;  
Amsterdam 370  
Montreal 5200  
Auckland 22720  
;
```

- **Column Input**

```
input city $ 1-11 distance 13-20;  
cards;  
Rome          1430  
Mexico City 10640  
Hong Kong    13200  
;
```

Lets look at the basic styles of input, Part 2

- Formatted Input

```
input @1 city $11. @13 distance 5.;  
cards;  
Stockholm      1450  
Los Angeles    8780  
Bangkok        12860  
;
```

- Delimited Input

```
infile cards dlm='~';  
input city $ distance;  
cards;  
Paris~330  
New York~5530  
Sydney~215660  
;
```

Lets look at the basic styles of input, Part 3

- "Double Space" Delimiter Input

```
input city $ & distance;  
cards;  
Frankfurt 640  
Rio de Janeiro 11060  
Singapore 10810  
;
```

- Named Input

```
input city= $ distance=;  
cards;  
city=Copenhagen distance=953  
distance=6800 city=Nairobi  
city=Tokyo distance=15260  
;
```

Combining two or more styles

DATA

2009-01-15	HYPERMAGNESEMA	MILD
2009-01-15	ABDOMINAL CRAMPS	MODERATE
2009-01-28	NAUSEA	MODERATE

INPUT STATEMENT

```
input startdt yymmdd10. event $ & severity $;
```

Named Styles to Transfer Data

```
1 data _null_;
2     set sashelp.class (obs=5);
3     put (_all_) (=);
4     run;
```

```
Name=Alfred Sex=M Age=14 Height=69 Weight=112.5
Name=Alice Sex=F Age=13 Height=56.5 Weight=84
Name=Barbara Sex=F Age=13 Height=65.3 Weight=98
Name=Carol Sex=F Age=14 Height=62.8 Weight=102.5
Name=Henry Sex=M Age=14 Height=63.5 Weight=102.5
```

Using this output and the Named Input style, it is possible to bring the data back into a dataset using the following code:

```
data class;
  length Name $8 Sex $1 Age Height Weight 8;
  input Name= $ Sex= $ Age= Height= Weight=;
cards;
Name=Alfred Sex=M Age=14 Height=69 Weight=112.5
Name=Alice Sex=F Age=13 Height=56.5 Weight=84
Name=Barbara Sex=F Age=13 Height=65.3 Weight=98
Name=Carol Sex=F Age=14 Height=62.8 Weight=102.5
Name=Henry Sex=M Age=14 Height=63.5 Weight=102.5
```

A More Complex Example

Cook and Eat (Yum Yum), Inc.
New Sauce

Page 1 of 19

Listing 4 Adverse Events

Subject Number: 1.001

```
-----  
#001  Event: HYPERMAGNESEMA  
       Preferred Term: Hypermagnesaemia  
       Body System: Metabolism and nutrition disorders  
       Start Date: 2009-01-15  Stop Date: 2009-01-25  
       Severity: MILD          Relationship: NONE          Action: NONE  
-----  
#002  Event: ABDOMINAL CRAMPS  
       Preferred Term: Abdominal pain  
       Body System: Gastrointestinal disorders  
       Start Date: 2009-01-15  Stop Date: 2009-01-27  
       Severity: MODERATE     Relationship: NONE          Action: NONE  
-----  
#003  Event: NAUSEA  
       Preferred Term: Nausea  
       Body System: Gastrointestinal disorders  
       Start Date: 2009-01-28  Stop Date: 2009-02-02  
       Severity: MODERATE     Relationship: POSSIBLE     Action: STUDY DRUG HELD  
-----
```

The Code

```
data ae_data;
  length ptnum $5 aenum $3 event $100 pt $100 soc $100
         startdt $10 stopdt $10 severity $10 relation $10 action $20;
  retain ptnum '';
  infile 'filename and location' length=len;
  input flg $varying200. len @;
  if flg=:'Subject Number:' then do;
    input @17 ptnum $5.;
  end;
  else if flg=:'#' then do;
    input @2 aenum $3. @14 event $ & /
         @23 pt $ & / @20 soc $ & /
         @19 startdt $ @42 stopdt $ /
         @17 severity $ & @43 relation $ & @63 action $ &;
    output;
    call missing(aenum,event,pt,soc,startdt,
                stopdt,severity,relation,action);
  end;
run;
```

Can everything be done?

Cook and Eat (Yum Yum), Inc.
New Sauce

Page 1 of 19

Listing 4 Adverse Events

Subject Number: 1.001

#001 Event: HYPERMAGNESEMA
Preferred Term: Hypermagnesaemia
Body System: Metabolism and nutrition disorders
Start Date: 2009-01-15 Stop Date: 2009-01-25
Severity: MILD Relationship: NONE Action: NONE

- Yes and no, but it is beyond this paper

Questions and Contact Information

Questions?

Contact Information

David Franklin

Independent SAS Consultant

16 Roberts Road, Litchfield, NH 03052

Phone: 603-275-6809

Email: dfranklinuk@compuserve.com

Web: <http://www.TheProgrammersCabin.com>

LinkedIn: <http://www.linkedin.com/in/davidfranklinnh>



Acknowledgement

SAS and all other SAS Institute Inc. product or service names are registered trademarks or trademarks of SAS Institute Inc. in the USA and other countries. ® indicates USA registration. Other brand and product names are registered trademarks or trademarks of their respective companies.