



Don't Be a SAS® Dinosaur: Modernize Your SAS Programs

by Warren Repole

Combine Data: Data Sets with Different Match Variables

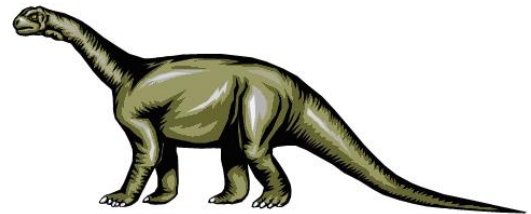
Scenario:

You need to match two data sets based on common values of one or more variables. Although the values in the two data sets match, the corresponding variable names do not match.

To match-merge data sets, the names of the BY variables must be the same in all data sets.

The old way: MERGE with RENAME= Option

The original approach is to (sort and) merge the data sets, applying the RENAME= data set option in order to align the variable name(s) for the match merge.



```
proc print data=prdtot2(obs=5);
  var MonYr Total2;
  title1 "PRDTOT2 data set";
run;
proc print data=prdtot3(obs=5);
  var Date Total3;
  title1 "PRDTOT3 data set";
run;

data merged;
  merge prdtot2(rename=(MonYr=Date)) prdtot3;
  by Date;
run;
proc print data=merged(obs=5);
  title1 "Results matching data sets with different key names";
run;
```

Please direct all correspondence to:

Warren Repole, 1705 Palm Springs Dr, Vienna VA 22182-2331

sasdinosaur@repole.com

www.repole.com/dinosaur

The new way: PROC SQL Join (available in SAS Version 6)

An alternate approach is to use an SQL query in which the WHERE clause can address the name differences explicitly.

```
proc print data=prdtot2(obs=5);
  var MonYr Total2;
  title1 "PRDTOT2 data set";
run;
proc print data=prdtot3(obs=5);
  var Date Total3;
  title1 "PRDTOT3 data set";
run;

proc sql;
  create table joined as
    select prdtot2.MonYr as Date, Total2, Total3
    from prdtot2, prdtot3
    where prdtot2.MonYr=prdtot3.Date;
quit;
proc print data=joined(obs=5);
  title1 "Results matching data sets with different key names";
run;
```

Advantages of the alternate approach:

- No RENAME= option syntax is required. This syntax typically is very tedious when multiple match variables are involved.
- No explicit sorting of the input data sets is required.

Disadvantages of the alternate approach:

- Additional data manipulation and programming logic is limited in an SQL query.
- The resources required by the SQL query often exceed the resources used by the combination of PROC SORT and the DATA step.

Additional documentation for this technique can be found in *SAS® 9.2 Language Reference: Dictionary*. Cary, NC: SAS Institute Inc.

Visit <http://support.sas.com/documentation/onlinedoc/sas9doc.html> for SAS 9 documentation.

Go to <http://www.repole.com/dinosaur/mergerename.html> for the sample code and output for this topic.

These techniques are mentioned in other SAS references and publications:

SAS Sample 25441 (<http://support.sas.com/kb/25/441.html>)

SUGI Paper 044-31 (<http://www2.sas.com/proceedings/sugi31/044-31.pdf>)