Web Products: The Marriage of PERL and QPL to Create SAS Code

Alice Feldesman, U.S. General Accounting Office, Washington, D.C.
Steven Palincsar, U.S. General Accounting Office, Washington, D.C.

ABSTRACT

A best practice in survey research is to display the questions and results of a survey. Web-based surveys allow rapid data collection, and they give readers of reports using Web-based surveys speedy access to the findings. The Web-based environment entices readers to seek multiple views of the data. The U.S. General Accounting Office (GAO) has developed its own software, QPL, for creating Web-based surveys. Our track record indicates that we may conduct about a hundred Web-based surveys each year. To meet the demand to display the results of the Web-based surveys, we created two Perl programs. One Perl program translates the QPL code into the SAS™ code, and the other organizes the html files that the respondents saw so that readers can readily access the results and navigate through the survey. This approach can create a highly accurate and realistic Web-based product in a few hours.

INTRODUCTION

We used three software programs to create Web-only products: QPL, SAS™ and Perl (a programming language).

Questionnaire Programming Language (QPL) is a software package that GAO developed. QPL describes questions. It creates many files to support Web-based surveys, among which are a pgm file and an html file of the survey. The pgm file is the file the survey designer creates that has the instructions, questions, response options, and navigation. This file has special QPL codes. The html file is layout respondents see when they answer the survey. While QPL begins with a single html file of the survey, deploying the survey involves generating numerous screens that a respondent navigates. In a Web-only product, each of these screens is a separate html file.

SAS™ software is also used to create Web-only products. SAS™ produces the calculations that appear in the tables, labels the tables, and controls the layout of the tables. SAS™ creates a table for each question. The output delivery system (ODS) and style sheets create the html table files.

Perl, a programming language, is very flexible in working with textual information. Staff in two GAO teams collaborated to create a Perl program that creates many files to support Web-based surveys, among which are a .an=, and ends with .next. This basic structure is uniform across all question types (check all that apply, text, matrixes).

As you can see, each question begins with .qu=, has a type=, has a.an=, and ends with .next. This basic structure is uniform across all question types (check all that apply, text, matrixes). Because of this uniformity, the code may be dismembered into its components and the results stored in a set of arrays and then repurposed into SAS™ code.

If you want to learn more about QPL, go to www.gao.gov/qpl. This site includes documentation and a tutorial, and in the near future it will contain a public release of the software. There is no charge for the software.

HOW GAO USED SAS™ TO CREATE WEB-BASED PRODUCTS IN THE PAST

Over the past several years, GAO has issued only a few Web-only products based on Web-based surveys. To create these products, an experienced SAS™ programmer wrote many lines of code to create the calculations and the tables to display the results. A second experienced SAS™ programmer reviewed the coding for technical accuracy. A variety of presentations were used, including lists with linked tables and a U.S. map with linked tables.

In 1999, in a report entitled Superfund: Half the Sites Have All Cleanup Remedies in Place or Completed (GAO/RCED-99-245), available on the Internet at http://www.gao.gov/cgi-bin/getrpt?RCED-99-245, a link was made to a list of states and a U.S. map. The perimeter and links from the U.S. map were manually digitized, using a mouse. Readers could access html tables that were generated by SAS™. As you will see, these tables are somewhat difficult to read because of their format, since color, font, alignment, and other formatting options were limited.

In 2000, a report concerning sprawl (Survey of Local Growth Issues RCED-00-272) was issued, available on the Internet at...
In 2001, a report on formerly used defense sites was issued. Entitled Environmental Contamination: Cleanup Actions at Formerly Used Defense Sites GAO-01-1012SP, it is available on the Internet at http://www.gao.gov/cgi-bin/getrpt?GAO-01-1012SP. The U.S. map and the links were generated through SAS™ code in this presentation. We added new features in this product, such as heading links for definitions.

While we successfully published these products on the Internet, this coding approach posed many challenges. A limited number of staff understood survey research, Web-based surveys, and Web-based products. Therefore, our flexibility on how many products could be supported and how quickly they could be produced and reviewed for accuracy was reduced. We had no written guidance to enable a range of staff to create a Web-based product that could be easily deployed. Guidance would have enabled staff to design their projects with standard naming conventions, style sheets, and layout. In addition, no mechanism ensured that when the main report was issued, the Web-only product was also posted on the Internet. Furthermore, when readers searched on the Internet, the searching convention would not necessarily return the two related products—the Web-based survey and the audit report it was based on.

To overcome these challenges, it was important to conceive an approach that could be supported by many SAS™ programmers and that could be sustained by the organizational structure within GAO. In addition, the approach had to be scaleable to meet the expected demand of a hundred Web-based surveys each year. While not all these surveys might produce a Web-only product, a more systematic approach was clearly needed.

**HOW PERL TRANSLATES QPL CODE INTO SAS™ CODE**

Just as QPL is a software package in which code is compiled, SAS™ code is compiled. This similarity is striking. When you write faulty code in QPL, you receive an error log, and files will not be created. Likewise, a log of errors is generated when your syntax is incorrect in SAS™, and files will not be created. Furthermore, to create an HTML table in SAS™ through PROC REPORT, for example, you will need the following: a data set, a set of fields, descriptors of fields (formats and headings), and a title. QPL, as noted above, contains these elements but stores them in the components of .qu, type, an, etc.

We looked to Perl to make the translation from QPL syntax into SAS™ syntax. Overall, in the translation Perl gives us labels, boiler plate macros, and invocation of the macros for each question (based on the type of question); it creates data sets to group common items (as in a matrix), a style sheet, and proc reports; and it tabulates for each question.

The advantage of Perl is that it can easily read text by paragraphs and the end-of-paragraph symbol can be defined as any arbitrary value. This lets you read a question as a single chunk. You can also decompose the paragraph elements, such as the question number, the type of question, and labels. These can be stored in an array. Perl lets you create a highly complex data structure that contains all the elements of the question in the order they appear in the survey.

Furthermore, having experience making Web-only products enabled us to troubleshoot syntax errors that an SAS™ programmer often has to debug. For example, the programmer often has to choose between single and double quotes in labels and titles. To meet this challenge, the Perl program examines the string to see if the survey designer used double quotes. If double quotes were used, then single quotes surround the text. Likewise, if single quotes were used, then double quotes surround the text. In this way, the problem of the missing quote is often avoided.

Another example of the value of using a systematic routine for producing the SAS™ code is the uniformity of the tables. Each table is produced by Perl. Consequently, the alignment, font, titles, and so on are always the same and also the syntax is never missing components. Writing code for hundreds of tables can lead to omission errors.

**AN EXAMPLE OF QPL CODE**

To help you understand this process, below is code from a GAO survey that has already been issued. It is only a few of the questions of the survey.

```
.qquestion=q1, type=mult
.ssubtitle=HIGHLY ERODIBLE LAND CONSERVATION (HELC) PROVISIONS"

During calendar years 1998-2001, how many status reviews on highly erodible land did your field office conduct?

.answer=0
.none
.goto Q21
1-10
11-20
21-30
31-40
41-50
More than 50
No response
.next

.qquestion=Q1a, type=mult
.In your opinion, compared with neighboring counties, how closely does your county monitor HELC provisions?

.answer=0
.Much more closely
.More closely
.About the same
.Less closely
.Much less closely
.Don't know
.next

.qquestion=q2, type=mult
.In your opinion, overall, what level of understanding do farmers in your county have about what constitutes a substantial reduction in soil erosion?

.answer=0
.Very great understanding
.Great understanding
.Moderate understanding
.Some understanding
.Little or no understanding
.Don't know
.next

.qquestion=q3, type=mult
.In your opinion, what level of understanding do farmers in your county have of
```
HELIC provisions?
.answer=q2
.next

.question=q4, type=mult
In your experience, to what extent are farmers in your county willing to comply with HELC provisions?
.answer=0
Very great extent
Great extent
Moderate extent
Some extent
Little or no extent
No response
.next

.question=q4a, type=mult
In your experience, how has farmers' willingness to comply with HELC provisions changed since 1996?
.answer=0
Much more willing to comply
More willing to comply
Neither more willing nor less willing to comply
Less willing to comply
Much less willing to comply
No response
.next

AN EXAMPLE OF SAS™ CODE RESULTING FROM THE PERL TRANSLATION

After the QPL survey is processed through Perl, the code that appears on the next several pages is generated. As you see, some pieces are missing, including the data source, any recoding of the data to ensure data quality, and the path where the files are stored. Fortunately, the QPL software generates these missing components. Staff simply cut and paste the missing pieces from the QPL software to complete the code. However, all the code below is generated by Perl and does not use a cut-and-paste approach. This eliminates errors associated with manually copying files.

OPTIONS NOSOURCE NOSTIMER NOCENTER;
data _null_;  
%PUT ****** --- ******* MM/DD/YYYY**; /* Replace information on line above as appropriate */  
%PUT *;  
%PUT * COMPUTER CODE VERIFICATION:
(Meets GAO policy to repeat calculation) *;  
%PUT *;  
%PUT * Calculations and code in this run were verified ______________ *;  
%PUT *;  
%PUT* on (date) ____/_____/_____ *;  
%PUT *;  

OPTIONS SOURCE;

/***********************************************************/

LABEL
q1 = "Q1. During calendar years 1998-2001, how many status reviews on highly erodible land did your field office conduct?"

q1a = "Q1a. In your experience, compared with neighboring counties, how closely does your county monitor HELC provisions?"

q2 = "Q2. In your opinion, overall, what level of understanding do farmers in your county have about what constitutes a substantial reduction in soil erosion?"

q3 = "Q3. In your opinion, what level of understanding do farmers in your county have of HELC provisions?"

q4 = "Q4. In your experience, to what extent are farmers in your county willing to comply with HELC provisions?"

q4a = "Q4a. In your experience, how has farmers' willingness to comply with HELC provisions changed since 1996?";
run;
/**************************************/
%macro tabmax (max,fd=fd);
proc freq data=fin noprint;
tables &fd/out=&fd;
where 1 <= &fd <= &max ;
run;
data &fd; set &fd end=eof;
DATA d3;
set &fd end=eof ;
retain
%do i = 1 %to &max ;
per%i. 0 ;
%end;
array per (&max.) per1-per&max. ;
%do i = 1 %to &max. ;
if &fd = &i. then per%i. = percent;
%end;
if eof then do ;
do i = 1 to &max. ;
if per(i) = . then per(i) = 0 ;
end; end;
keep %do i = 1 %to &max ;
per&i.  %end;;
run;
data d3; set d3 end=eof;
if eof then output; run ;
proc transpose data=&fd out=&fd; run;
DATA d2;
SET &fd (WHERE=(\_NAME_ EQ 'COUNT' ));
%do i = 1 %to &max . ;
if col&i. = . then col&i. = 0 ;
%end;
ncases=sum (of col1-col&max );
keep ncases; run;
data d1; length _LABEL_ $350. ;
SET &fd
(WHERE=(\_NAME_ NE 'COUNT' AND \_NAME_ NE 'PERCENT' ));
keep _LABEL_; run;
DATA &fd; MERGE d1 d2 d3;
run;
PROC DATASETS LIBRARY=WORK NOLIST;
delete d1 d2 d3;
run;
%mend tabmax;
data fin; set fin;
run;
%tabmax (7, fd=q1); run;
%tabmax (5, fd=q1a); run;
%tabmax (5, fd=q2); run;
%tabmax (5, fd=q3); run;
%tabmax (5, fd=q4); run;
%tabmax (5, fd=q4a); run;
/* You don't want "asis=on" if you are building a product for web use. This code is useful only for conversion to a print pub. It adds <PRE></PRE> tags around the titles, which are in table cells. */
options nonumber nodate; run;
ods listing close;
proc template;
define style styles.test;
parent=styles.minimal;
style systemtitle from systemtitle /
font_face=helvetica font_size=3
asis=on;
style systemfooter from systemfooter /
font_face=helvetica font_size=3;
style header from header/
font_face=helvetica
cellwidth=1 in
just=r vjust=b;
style data from data/
font_face=helvetica;
style rowheader from rowheader/
font_face=helvetica;
style body from body/
pagebreakhtml=_undef_;
run;
options orientation=landscape;
/*****************************/
ods html
body="PUT_YOUR_PATH_HERE.html"
(title='U.S. General Accounting Office') style=styles.test;
/*****************************/
/* Normal mult question, 7 columns */
/* Proc report for mult question q1 */
/* NOTE: if you repeat your scale you will get two or more proc reports for the same matrix data set */
proc report data=q1 nowindows spacing=1
pspace=1 split='*' missing headline;
columns per1 per2 per3 per4 per5 per6 per7 ncases;

define per1 /display format=4.1 'None*(percent)' style=[cellwidth=1 in]  
style=header [just=right];
define per2 /display format=4.1 '1-10*(percent)' style=[cellwidth=1 in]  
style=header [just=right];
define per3 /display format=4.1 '11-20*(percent)' style=[cellwidth=1 in]  
style=header [just=right];
define per4 /display format=4.1 '21-30*(percent)' style=[cellwidth=1 in]  
style=header [just=right];
define per5 /display format=4.1 '31-40*(percent)' style=[cellwidth=1 in]  
style=header [just=right];
define per6 /display format=4.1 '41-50*(percent)' style=[cellwidth=1 in]  
style=header [just=right];
define per7 /display format=4.1 'More than 50*(percent)' style=[cellwidth=1 in]  
style=header [just=right];
define ncases/display format=comma5.0 'Number*of respondents'  
style=[cellwidth=1 in] style=header [just=right];
TITLE "Q1. During calendar years 1998-2001, how many status reviews on highly  
erodible land did your field office conduct?";
runt;
/* Normal mult question, 5 columns */
/* Proc report for mult question q1a */
/* NOTE: if you repeat your scale you will get two or more proc reports for  
the same matrix data set */

proc report data=q1a nowindows spacing=1
pspace=1 split='*' missing headline;
columns per1 per2 per3 per4 per5 ncases;

define per1 /display format=4.1 'Much more closely*(percent)'  
style=[cellwidth=1 in] style=header [just=right];
define per2 /display format=4.1 'More closely*(percent)'  
style=[cellwidth=1 in] style=header [just=right];
define per3 /display format=4.1 'About the same*(percent)'  
style=[cellwidth=1 in] style=header [just=right];
define per4 /display format=4.1 'Less closely*(percent)'  
style=[cellwidth=1 in] style=header [just=right];
define per5 /display format=4.1 'Much less closely*(percent)'  
style=[cellwidth=1 in] style=header [just=right];
define ncases/display format=comma5.0  
'Number*of respondents' style=[cellwidth=1 in] style=header [just=right];
TITLE "Q1a. In your experience, compared with neighboring  
counties, how closely does your county monitor HELC provisions?";
runt;
/* Normal mult question, 5 columns */
/* Proc report for mult question q2 */
/* NOTE: if you repeat your scale you will get two or more proc reports for  
the same matrix data set */
proc report data=q2 nowindows spacing=1
pspace=1 split='*' missing headline;
columns per1 per2 per3 per4 per5 ncases;

define per1 /display format=4.1 'Very great understanding*(percent)'  
style=[cellwidth=1 in] style=header [just=right];
define per2 /display format=4.1 'Great understanding*(percent)'  
style=[cellwidth=1 in] style=header [just=right];
define per3 /display format=4.1 'Moderate understanding*(percent)'
    style=[cellwidth=1 in] style=header
    [just=right];

define per4 /display format=4.1 'Some understanding*(percent)'
    style=[cellwidth=1 in] style=header
    [just=right];

define per5 /display format=4.1 'Little or no understanding*(percent)'
    style=[cellwidth=1 in] style=header
    [just=right];

define ncases/display format=comma5.0 'Number*of respondents'
    style=[cellwidth=1 in] style=header
    [just=right];

TITLE "Q2. In your opinion, overall, what level of understanding do farmers
in your county have about what constitutes a substantial reduction in
soil erosion?";
run;

/* Normal mult question, 5 columns */
/* Proc report for mult question q3 */
/* NOTE: if you repeat your scale you
will get two or more proc reports for
the same matrix data set */

proc report data=q3 nowindows spacing=1
    pspace=1 split='*' missing headline;
columns per1 per2 per3 per4 per5
ncases;

define per1 /display format=4.1 'Very great understanding*(percent)'
    style=[cellwidth=1 in] style=header
    [just=right];

define per2 /display format=4.1 'Great extent*(percent)'
    style=[cellwidth=1 in] style=header
    [just=right];

define per3 /display format=4.1 'Moderate extent*(percent)'
    style=[cellwidth=1 in] style=header
    [just=right];

define per4 /display format=4.1 'Some extent*(percent)'
    style=[cellwidth=1 in] style=header
    [just=right];

define per5 /display format=4.1 'Little or no extent*(percent)'
    style=[cellwidth=1 in] style=header
    [just=right];

define ncases/display format=comma5.0 'Number*of respondents'
    style=[cellwidth=1 in] style=header
    [just=right];

TITLE "Q3. In your opinion, what level
of understanding do farmers in your
county have of HELC provisions?";
run;

/* Normal mult question, 5 columns */
/* Proc report for mult question q4 */
/* NOTE: if you repeat your scale you
will get two or more proc reports for
the same matrix data set */

proc report data=q4 nowindows spacing=1
    pspace=1 split='*' missing headline;
columns per1 per2 per3 per4 per5
ncases;

define per1 /display format=4.1 'Very great extent*(percent)'
    style=[cellwidth=1 in] style=header
    [just=right];

define per2 /display format=4.1 'Great extent*(percent)'
    style=[cellwidth=1 in] style=header
    [just=right];

define per3 /display format=4.1 'Moderate extent*(percent)'
    style=[cellwidth=1 in] style=header
    [just=right];

define per4 /display format=4.1 'Some extent*(percent)'
    style=[cellwidth=1 in] style=header
    [just=right];

define per5 /display format=4.1 'Little or no extent*(percent)'
    style=[cellwidth=1 in] style=header
    [just=right];

define ncases/display format=comma5.0 'Number*of respondents'
    style=[cellwidth=1 in] style=header
    [just=right];

TITLE "Q4. In your experience, to what
extent are farmers in your county
willing to comply with HELC
provisions?";
run;

/* Normal mult question, 5 columns */
/* Proc report for multi question q4a */
/* NOTE: if you repeat your scale you will get two or more proc reports for the same matrix data set */

```plaintext
proc report data=q4a nowindows
spacing=1 pspace=1 split='*' missing headline;

columns per1 per2 per3 per4 per5 ncases;

declare per1 /display format=4.1 'Much more willing to comply*(percent)'
style=[cellwidth=1 in] style=header [just=right];

declare per2 /display format=4.1 'More willing to comply*(percent)'
style=[cellwidth=1 in] style=header [just=right];

declare per3 /display format=4.1 'Neither more nor less willing to comply*(percent)'
style=[cellwidth=1 in] style=header [just=right];

declare per4 /display format=4.1 'Less willing to comply*(percent)'
style=[cellwidth=1 in] style=header [just=right];

declare per5 /display format=4.1 'Much less willing to comply*(percent)'
style=[cellwidth=1 in] style=header [just=right];

declare ncases/display format=comma5.0 'Number*of respondents'
style=[cellwidth=1 in] style=header [just=right];

TITLE "Q4a. In your experience, how has farmers' willingness to comply with HELC provisions changed since 1996?";
run;

ods html close; run;
```

/* This code makes the web only version with a file for each question. However, you will need to add PROC FORMAT and a format statement which you can get from the QPL system. One correction will need to be made to the value statements: you will need to add "(percent)" to each of items in the scale. */

```plaintext
proc template;
define style styles.test2;
parent=styles.minimal;

style systemtitle from systemtitle /
font_face=helvetica font_size=3
font_weight=bold;

style systemfooter from systemfooter /
font_face=helvetica font_size=3;

style header from header/
font_face=helvetica
cellwidth= 1 in
just=r vjust=b;

style data from data/
font_face=helvetica;

style body from body/
pagebreakhtml=_undef_;

style rowheader from rowheader/
font_face=helvetica;

end;
run;
ods listing close;
```

```plaintext
data fin; set fin;
row=/* SUBSTITUTE YOUR ROW VARIABLE HERE */;
run;

%macro tab (fd=fd,title=title);
ods html body="/FULLPATH_GOES_HERE&fd..html" (title='U.S. General Accounting Office') style=styles.test2
headtext="<style>
p{display:none}</style>";
proc tabulate data=fin missing;
class row &fd/preloadfmt exclusive;
tables row=' ' all='/* PUT YOUR ROW TITLE HERE */',
(&fd=' '*pctn<&fd>=' '*f=10.1 n='Number of respondents' f=comma8.)/indent=2
rts=15 misstext='0' printmiss;

Title 'U.S. General Accounting Office';
title2 'Survey on xxxxxxxxx (GAO-0x-xxxxSP)';

Title3 &title; run;
%mend tab; run;
ods html close; run;
```
/* You need one of these for each question that has a table with means, medians, etc
If your row variable is already formatted then in the macro eliminate the ", format=format"
and in the invocation eliminate the ", format=" */

%macro tabmean (fd=fd,title=title,
format=format);
proc sort data=fin; by row;
ods listing;
PROC UNIVARIATE NOPRINT DATA=FIN;
by row; VAR &fd;
OUTPUT OUT=FIN&fd N=NVAL MEAN=MEANVAL 
MEDIAN=MEDIANVAL MIN=MINVAL MAX=MAXVAL;
proc freq data=fin noprint; tables &fd/out=lab&fd; proc transpose
data=lab&fd out=lab&fd; run;
data &fd; SET lab&fd (where= (_NAME_ NE 'COUNT' AND _NAME_ NE 'PERCENT'));
keep _LABEL_; run ;
data all&fd; merge fin&fd &fd; run;
PROC UNIVARIATE NOPRINT DATA=FIN;
VAR &fd; OUTPUT OUT=FIN&fd N=NVAL MEAN=MEANVAL MEDIAN=MEDIANVAL
MIN=MINVAL MAX=MAXVAL;
proc freq data=fin noprint; tables &fd/out=lab&fd; proc transpose
data=lab&fd out=lab&fd; run;
data &fd; SET lab&fd
(where= (_NAME_ NE 'COUNT' AND _NAME_ NE 'PERCENT'));keep _LABEL_; run ;
data all2&fd; merge fin&fd &fd; run;
data all&fd; set all2&fd;
ods listing close;
ods html body="/"FULLPATH_GOES_HERE
"/fd..html" (title='U.S. General Accounting Office') style=styles.test2
headtext="cstyle > p<{display:none}</cstyle>";
proc report data=all&fd nowindows
spacing=1 pspace=1 split='*' missing
headline;
columns row MEANVAL MEDIANVAL MINVAL MAXVAL NVAL;
define MEANVAL /display format=4.1
'Mean' style (header)=\{just=r\};
define MINVAL /display format=4.0
'Minimum' style (header)=\{just=r\};
define MAXVAL /display format=4.0
'Maximum' style (header)=\{just=r\};
define MEDIANVAL /display format=4.0
'Median' style (header)=\{just=r\};
define NVAL/display format=comma5.0
'Number*of respondents' style
(header)=\{just=r\};
Title 'U.S. General Accounting Office';
title2 'Survey on xxxxxxxxxxxxx (GAO-03-xxxxSP)';
title3 &title; run;
%mend tabmean;
r
%tab (fd=q1, title="Q1. During calendar years 1998-2001, how many status reviews on highly erodible land did your field office conduct?"); run;

%tab (fd=q1a, title="Q1a. In your experience, compared with neighboring counties, how closely does your county monitor HELC provisions?"); run;

%tab (fd=q2, title="Q2. In your opinion, overall, what level of understanding do farmers in your county have about what constitutes a substantial reduction in soil erosion?"); run;

%tab (fd=q3, title="Q3. In your opinion, what level of understanding do farmers in your county have of HELC provisions?"); run;

%tab (fd=q4, title="Q4. In your experience, to what extent are farmers in your county willing to comply with HELC provisions?"); run;

%tab (fd=q4a, title="Q4a. In your experience, how has farmers' willingness to comply with HELC provisions changed since 1996?"); run;

As you can see a great deal of SAS™ code is generated through the Perl program. This is the code for eight questions. If you would like to view the entire survey associated with this project, it can be found in the report Agricultural Conservation: Survey Results on USDA's Implementation of Food Security Act Compliance Provisions (GAO-03-492SP), available on the Internet at http://www.gao.gov/cgi-bin/getrpt?gao-03-492SP. Here is a sample page from the survey.
GAO analysts must perform some edits before their Web-based products are ready for posting. We provided comments throughout the SAS™ program. Key notes include:

Add a PROC FORMAT called com (we provide the code as a comment). This will enable you to have format for any narrative questions and you want the table to display the percentage of respondents who provided a comment. The percentage is based on a person writing one character. You may want to change this threshold to five or more characters, depending on your project.

Enter the path where you would like SAS™ to write the tables. You will find this note where ODS appears in the program.

If a table has more than seven columns, we recode your data so that the resulting table is vertical rather than horizontal.

For some projects, teams would like tables that show subgroups such as states. The SAS™ program is written generically, using the variables name row to represent the subgroup. You will need to recode your subgroup variable to the field name row.

Sometimes a survey question is very long. This often happens when instructions are embedded in the question. To avoid compilation errors, we truncate all questions longer than 256 characters. We include a warning comment when this happens. You will need to edit the labels and titles for these questions, as appropriate.

**HOW PERL TRANSLATES THE HTML SURVEY INTO THE WEB-ONLY PRODUCT**

We wrote a second Perl program to take the single html file that was posted on a Web server to create the screen shots that the respondent viewed. Through Perl, this presentation provided the table of contents, the navigation across screens, and the links to each html table (that was generated by SAS™). By giving the files generated by the SAS™ program the same name as the question number and making a link with the same name, all the pieces fit together.

**HOW GAO USES SAS™ TO CREATE WEB-BASED PRODUCTS TODAY**

We now have guidance for staff to hand off the QPL files to analysts so that Web-based products can be produced in a few hours. We have four Web-based surveys that will be using this technology in summer 2003 and that should be available on GAO’s Web site later in the year. If you are wondering why it takes so long to produce these Web-based products, it is that each one accompanies an analytical report, another GAO product that is the result of many months of audit work performed by GAO staff and that is issued at the same time as the Web-only presentation of the Web-based survey results.

One enhancement we are considering is adding “fly-overs” or alternative tags on the tables. When readers view an html table, the heading often scrolls off the screen. A fly-over is a small box that appears when a mouse is held over a cell in a table. This enhancement will enable readers to easily understand the row and column heading by displaying this information in an alternative tag.

Another future step is the recognition that a Web-only product can be produced through this technology without needing to deploy a Web-based survey. When we have a data set that might benefit from multiple views of the data on the Internet, we can write a “mock” Web-based survey and use the Perl program to generate the SAS™ code. This code makes the multiple views of the data that can be posted on the Internet. We expect to have such a presentation on the Internet in early summer 2003. We see enormous potential in this alternative Web-only product.

Finally, we are pursuing with management the notion that Web-based survey presentations should be a new product line for the agency, since they are Web-based only. This product does not follow any of GAO’s conventional steps for producing its written products. In order for search engines to successfully identify these surveys, it would be advantageous for the agency to develop them as a unique product line. Moreover, the level of effort necessary to create a Web-based survey presentation requires only a few hours, not several days. Therefore, it would be inappropriate to “count” these presentations in our performance indicators.

**CONCLUSION**

Presenting a survey and its associated summary statistics is a best practice in survey research. In fact, the American Association of Public Opinion Research (AAPOR), a professional association whose members include many survey researchers, has a minimal standard of disclosing “The exact wording of questions asked, including the text of any preceding instructions or explanation to the interviewer or respondents that might reasonably be expected to affect the response.” But the effort to develop presentations has been so enormous that often, at best, the questions are merely reproduced and the results tabulated. Time-consuming word processing and verification are necessary to ensure the information’s accuracy, and this approach fails to meet agency core values of transparency. Researchers cannot see how the instrument was presented to respondents, and respondents are often unable to know how their responses are used. The ability to replicate surveys is reduced. This inaccessibility may reduce respondents’ willingness to complete surveys in a timely manner or researchers’ ability to achieve high response rates.

Our innovation is in writing Perl programs as a translator for presenting the survey instrument in the form that the respondent sees, as well as the summary table results. This rapid deployment provides an attractive and realistic presentation of survey results. Since surveys provide a critical source of information for policy makers, it is very important that surveys continue to be conducted and their results be made available to all American citizens.

**REFERENCES**

CONTACT INFORMATION

Your comments and questions are valued and encouraged.
Contact the author at
Alice Feldesman
U.S. General Accounting Office
441 G Street, NW, Room 6C22
Washington, DC  20548
Work Phone: 202-512-4927
Fax: 202-512-6171
E-mail: Feldesmana@gao.gov

Steve Palincsar
U.S. General Accounting Office
441 G Street, NW, Room 1825M
Washington, DC  20548
Work Phone: 202-512-3947
E-mail: PalincsarS@gao.gov

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.