Paper #-AD12

Bridging the Gap: From Spreadsheets to Websites

Maryanne Hope, Destiny Corporation
Mary K. Tucker, Royal & SunAlliance Insurance Co., Charlotte, NC

ABSTRACT

Business needs come in all shapes and sizes and computer technology continues to improve in leaps and bounds. Combine these conditions with savvy business users and the result is new demand for on-line Internet information systems solutions. And quick! Following is an account of one such application development solution. Using SAS® tools such as SAS/Access to PC files, SAS Intrnet and others, the authors combined efforts to provide an Intranet based on-line look up system, web enabling data from 56 sources and giving the business user a head start on the future. We will provide a step by step look at the initial conditions, the development process and the resulting application.

INTRODUCTION

Technology solutions are developed to meet business needs. The onset of the information super highway, and especially the internet, have sparked business users' imaginations and given them new ideas about how to better meet the demands of their jobs on a daily basis.

In our case, the business user has spent much of his career as an engineer, specializing in boiler and machinery inspection. He and his county-wide team are dedicated to improving the safety of their clients and communities at large by ensuring corporate machinery is working properly and offering additional safety solutions during annual on-site inspections.

Through the years as a working manager he developed ideas about technology solutions that would make his work and the work of his colleagues more efficient and timely.

THE BUSINESS CASE

Boiler and machinery inspection in the United States is heavily regulated at the state level and sometimes even at the municipal level. Each of the 56 jurisdictions has its own set of rules and regulations about the inspection process and resulting data transmission and collection. Through the Electronic Freedom of Information Act Amendments of 1996 (E-FOIA), the jurisdictions are obligated to share the information collected about jurisdictional objects and inspections at the site (or address) level upon request. They are, however, free to choose the medium and format through which to provide this information, leading to essentially 56 different data formats, one for each jurisdiction, in various spreadsheet, flat file and other formats.

Nonetheless, our savvy business user knew there must be a way he could consolidate this information across all jurisdictions and make it available to his constituents quickly and efficiently without a large investment in capital and resources. Certain the benefits would include positive financial impact on the corporation through improved planning and scheduling of inspections and more timely management of the fees and costs associated with site engineering inspections, he was open to suggestions and options.

IS THERE A SOLUTION?

After years of contemplation, dedication and research our user and others like him came to realize there is no easy or simple solution.

The best solutions include wide scale implementation across jurisdictions and inspection engineers to standardize many aspects of the inspection process, data collection and reporting. And in fact, a consolidated effort is under way.

But that didn't solve our client's problem for today. So we began to entertain the idea of using SAS as a short to mid-term solution. Our existing production SAS environment already included all the technology necessary to:

- import spreadsheets (using SAS/Access to PC Files)
- transform, cleanse and integrate data (using Base SAS)
- and provide web enabled access to data for reporting purposes (using SAS/Intrnet)

Given this environment, there was only one piece of the equation left: a developmental partner with the expertise and manpower available to do the work. SAS Quality Partner, Destiny Corporation, provided the missing piece to the puzzle.

THE DEVELOPMENT PROCESS: DATA STANDARDIZATION AND STRUCTURE

Destiny Corporation professionals were quick to point out "It's not the first time we have experienced data received from here, there and everywhere and no standardization to be found!" So, in implementing the first phase of this project they realized they were starting a challenge, as the incoming data for the jurisdictional database would be received from approximately 56 sources.

The processes identified to begin were identified were to standardize and clean the data in order to build a comprehensive database for access by a front end WEB query application. The data was received in a variety of file formats including text, EXCEL, WORD, ACCESS and CSV. It was discovered early on that complete automation of this stage was not possible because of the non-standard field names and variability of the fields collected from each jurisdiction. These variant columns had to be converted and mapped to the twenty-four standard fields identified for the query database.

The challenge was that the data coming in did not always match the fields, the naming conventions or the required attributes for the standardized structure. As a preliminary step, Proc Import was used to convert all the incoming files to a SAS data file.

Once in a SAS format, Proc Contents was run on the dataset to determine what fields were present and available for mapping to the standard fields. The next step was a manual process of reviewing the data for content (yes, there still is the need to eyeball the data!) coupled with running frequencies (PROC FREQ) to determine exactly what fields were present and their values. Detective work was necessary in order to decipher and map the values to the appropriate fields. This proved exceptionally difficult when column headings were not provided.

A 'standardizing' SAS program was written to convert the 'discovered' fields to the standard fields with the required attributes. Below is an example of some of the SAS functions used to convert to standard variables.

```
state_number=        put(inspequ_statenumbr,$12.) ;

manufacturer=        upcase(compress(put(manufname,$25.),'*'));

boiler_type=
upcase(compress(put(typeofequi,$35.),'/'));

owner_name=          upcase(put(ownername,$45.));

last_inspected=      datepart(inspecdate);

certificate_expiration= input(c_expire_dte,mmddyy8.);
```

Convenient SAS functions like 'PUT' which store the value in the specified format and 'COMPRESS' which remove extraneous characters in the text fields were extremely useful in the conversion. The 'UPCASE" function converted all character strings to upper case for text standardization. The format for date fields is 'mmddyy9.' but often the dates were received in a date-time format. To strip off the time portion the 'DATEPART' function was used to extract month/day/year values.

After standardization of the structure, the next step was to "cleanse." (The imported data ranged from "squeaky clean" to "down right dirty!") The process of cleansing the data included stripping out unwanted characters, correcting spellings, standardizing variations of character strings and eliminating unwanted delimiters or comments by the data entry personnel. This step was handled by writing customized SAS code addressing the specific issues in each of the jurisdictions.

In addition to SAS programs, commercial software MAILER 4+ was used to supply missing address, city, state and zip code data from an internal lookup database.

The final step was to build the comprehensive jurisdictional database and to create indexes on the key fields to enhance query performance. This was achieved by appending each "scrubbed 'individual jurisdictional data set" to the master data set using PROC APPEND for appending and PROC DATASET for indices.

To supply data driven values for the pull down selections on the appropriate query fields, a lookup database was created using PROC FREQ. All distinct values are discerned and contained in the frequency table for the key query fields.

SAS WEB APPLICATION RESULTS

At last, the database is ready for querying! Now on to constructing a straight forward, intuitive front end for users to find answers to their business needs.

The task at hand is to build a query system accessible to many individuals in the organization which is intuitive to the user while providing immediate results with reliable information. SAS/IntrNet technology provided just this platform. The application enables the user to access the jurisdictional database without having to know SAS, is available on the Intranet system in the company allowing multiple individuals simultaneous querying, and eliminates the need for client-side plug-in or installed software. The design of the application allows the user to construct simple queries via the WEB server against the comprehensive database and report the results to the user in the format of choice.

A brief description at a high level view of the SAS Web Application development phase follows. The components required for the SAS/IntrNet software to function are:

- a WEB Server such as Microsoft /IIS or Apache *
- a WEB Browser such as MSN or Netscape
- and components of the SAS IntrNet Software including the SAS Application Server and the Application Broker (CGI).

A Pool service was set up to provide a 'pool' of application servers shared by the clients.

The self-generated WEB pages were created by embedding the HTML code in the SAS code. The powerful Data Step was used to insert tags, parameters and attributes. The value of pull down fields or data-entry input from the WEB pages are passed as macro variables to the SAS code for processing.

The first screen (Figure 1) allows users to select a state or multi-select combination of states to query jurisdictional object data.

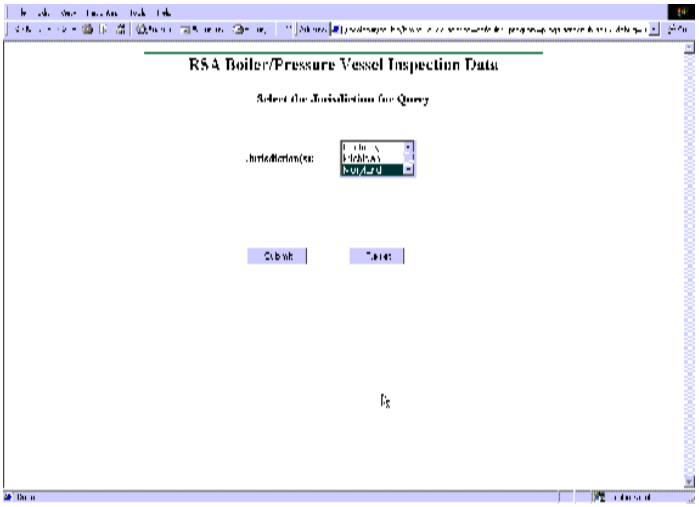
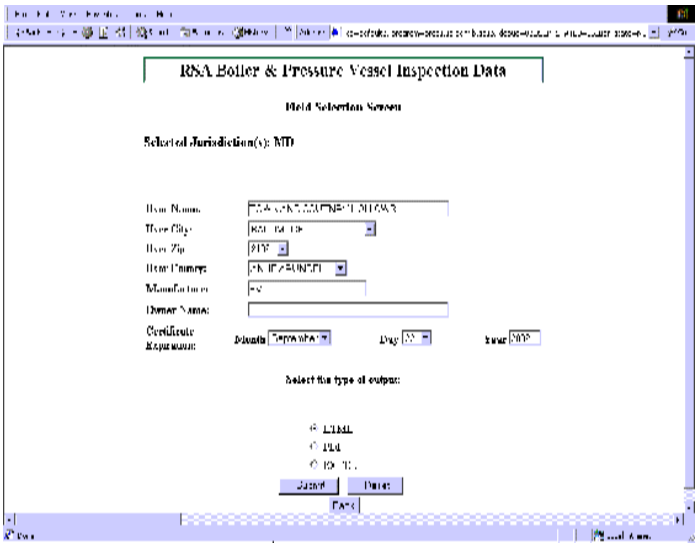


FIGURE 1

Single or multiple selections for the state field are permitted. The SAS running behind the scenes generates the macro variable for the 'state' field, which is then



passed to the next screen (SAS program) subsetting the data. Submit brings the user to the second screen (FIGURE 2).

FIGURE 2.

From this screen the user has the capability of selecting a single field or multiple fields to query the data. All character strings are converted to upper case. The fields with pull down selections were originally captured in the frequency cross tabulation. The decision to create a certain type of input as a pull down or user-typed field was determined by the possible number of distinct values in the data.

The screen was developed using standard HTML code in the SAS program. Each field selection generates a ‘where also’ condition on the data to populate the output dataset. On this screen the user was provided with a choice of output formats: HTML, PDF, or EXCEL.

The ability to store the output provides the potential of sharing information via email as an attachment, creating a hard copy report or generating an EXCEL spreadsheet for further manipulation. If the query does not match any values in the database, feedback to the users is given in the form of a message displayed on the screen. The last screen is the result of the users field(s) selection and the choice of output. (FIGURE 3)

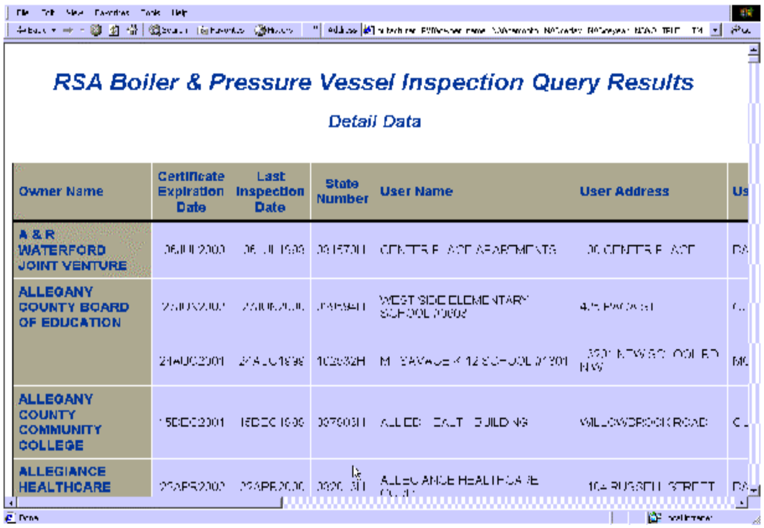


FIGURE 3.
This sample output shows the query result in HTML format. The output is generated by using a simple PROC PRINT with ODS statements. Some snippets of the SAS code shows the syntax written for each type of output offered:

```
If &output HTML then...
  Odds html file=_web out (dynamic) style=default;
         Proc print......
    Data _null_;
     File _web out;
     ...
        ...
    Run;
  ods html close;

Else if &output PDF then ....
      data _null_;
         rc=appsrv_header('Content-type:',
                          'application/pdf');
      run;
      ods printer PDF  notoc file=_webout style=default;
               Proc Print .....
      ods printer PDF close;

 Else %if &output = EXCEL ....
    %ds2csv(data=work.subset,csvfref=_webout);
```

ACKNOWLEDGMENTS
The authors offer a special thanks to Jerry Sturch at Royal & SunAlliance for his vision, passion and patience.

CONTACT INFORMATION
Your comments and questions are valued and encouraged. Contact the author at:

Maryanne Hope
Destiny Corporation
100 Great Meadow Rd.
Suite 601
Wethersfield, CT 06109
860.721.1684 ext. 16
mhope@destinycorp.com
www.destinycorp.com

Mary K. Tucker
Royal & SunAlliance Insurance Company
9300 Arrowpoint Blvd.
Charlotte, NC 28201
704.522.0348
Mary_Tucker@rsausa.com