

## Abstract

For large surveys, creating a comprehensive codebook can present many challenges. Without automation, this process can be labor intensive and error-prone; data in the codebook can be outdated or inconsistent with the actual data in the datasets. Another major challenge is that information/data for codebooks can come from multiple sources, such as questionnaire specifications, the questionnaire design system, and database or SAS datasets containing collected data. Our poster presents an integrated approach to codebook generation using modern tools and technologies, including SAS dictionary tables and SAS Integrated Object Model (IOM) for data management, HTML/CSS for codebook presentation, and the .NET framework for integrating the pieces together.

## 1. Introduction

### Assumptions:

The SAS dataset from which the codebook will be created from has gone through all data processing required for the project and is ready for delivery. As such, it should contain formats and labels for all variables in the dataset. Additional survey metadata, such as question text, special notes, imputation formulation, and other information, may also be available for inclusion with the codebook.

### Technologies:

The following technologies were used to leverage the strengths in each technology:

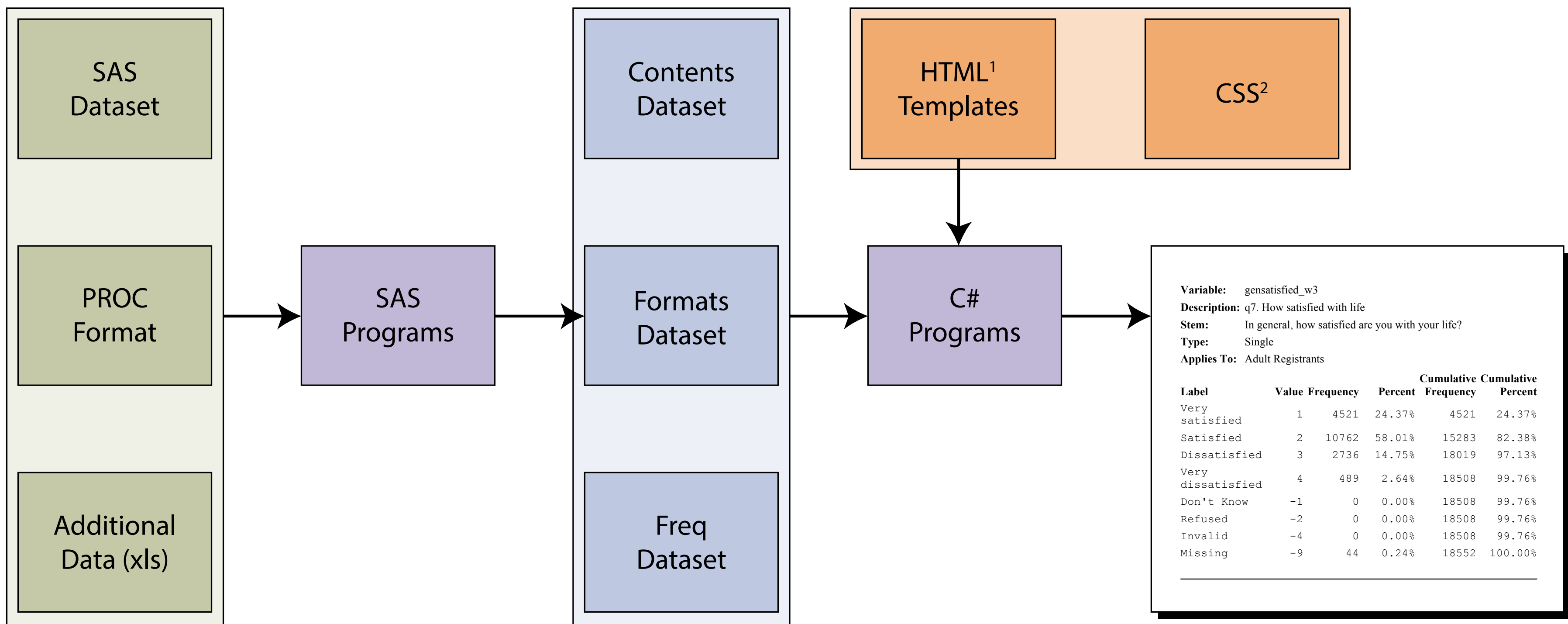
- SAS® Software– Statistical analysis, handling large amounts of data or computation-intensive applications
- .NET Framework – Microsoft’s platform for building windows and web applications
- SAS Integrated Object Model (IOM) – SAS IOM provides a bridge between SAS and .NET platforms
- HTML/CSS – Standardized styling and structure of data

### Implementation:

The system has 2 components:

1. SAS programs are responsible for reading the input SAS datasets and creating a set of SAS datasets:
  - Contents dataset – Contains a list of variables in the input dataset and their attributes
  - Formats dataset – Contains all format definitions that will be used in the codebook
  - Frequencies dataset – Contains the frequency table for each variable
  - Survey Metadata dataset – Contains additional metadata specific to a survey
2. .NET program utilizes data from codebook datasets and a set of predefined HTML templates to produce the codebook. The key features of this program are:
  - Reads SAS datasets using SAS IOM and .NET ADO.NET technologies. Data access is done in SQL.
  - Uses HTML to provide document structure.
  - Uses CSS to format and style codebook.
  - Changes the presentation of the codebook without having to change programming codes by using HTML template and CSS.

## 2. Data Flow Diagram



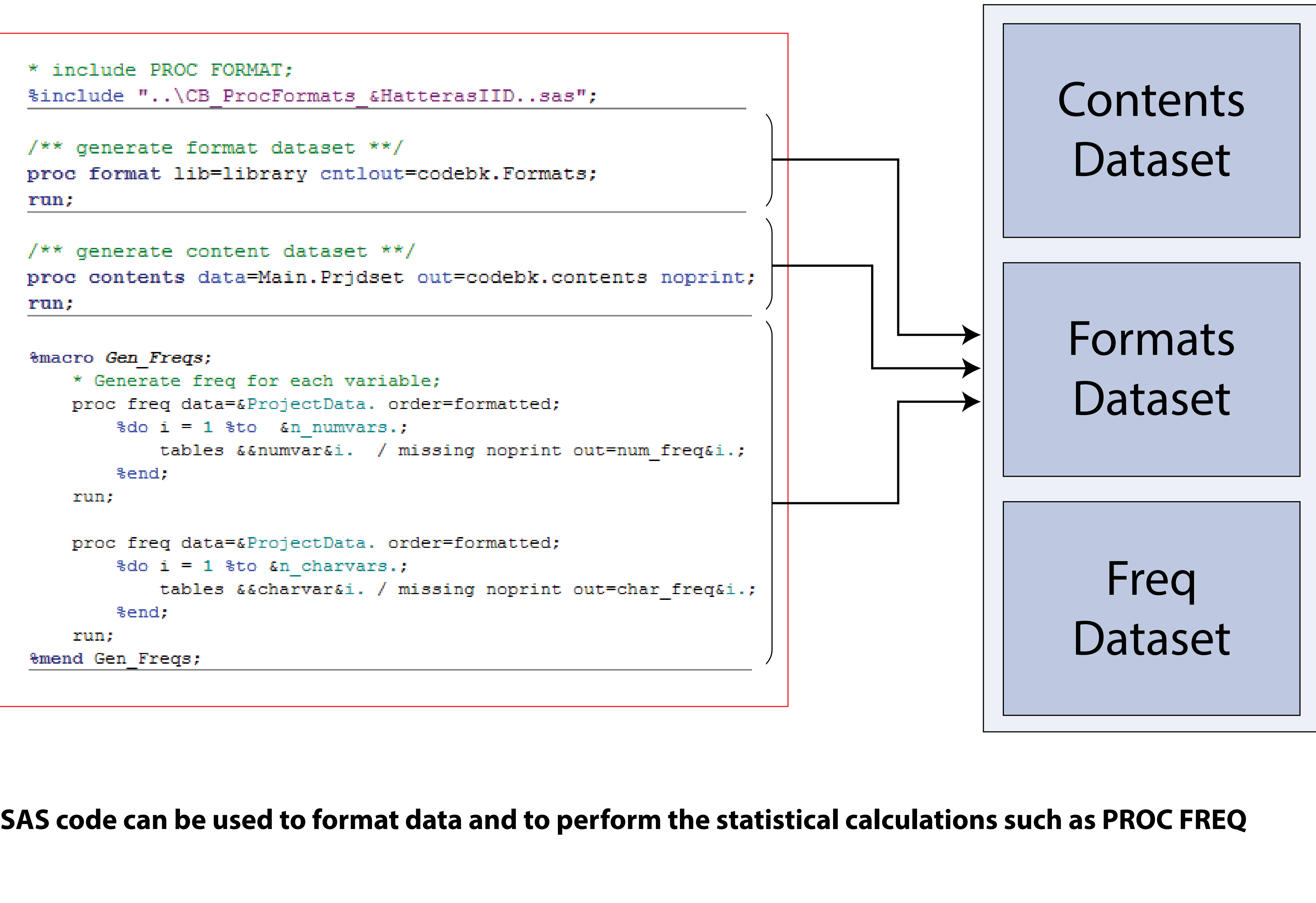
**SAS® software launches the codebook process and then the resulting template is stylized and saved using C#**

<sup>1</sup> HTML – Hypertext Markup Language, <http://www.w3.org/community/webed/wiki/HTML>

<sup>2</sup> CSS – Cascading Style Sheet, <http://www.w3.org/Style/CSS/Overview.en.html>

## 3. Codebook Data Gathering with SAS

Code snippets to generate codebook datasets from the SAS programs are shown below. The SAS programs are responsible for generating codebook data: labels, formats, frequencies, and other notes. These tasks were done using standard SAS procedures: PROC CONTENTS, PROC FORMAT, PROC FREQ, and PROC SQL. Codebook data are stored in SAS dataset formats, which are accessible to the .NET program via SAS IOM.



**SAS code can be used to format data and to perform the statistical calculations such as PROC FREQ**

## 4. Templates with HTML/CSS

Several examples of the HTML templates and the CSS styles are shown below. These templates and CSS provide several benefits:

- Codebook structure can be easily manipulated without code changes
- Codebook layout, font selection, color, and styles can be easily controlled and adapted to the requirements
- CSS is easy to change and requires nothing more than a text editor like Notepad
- Provides separation of style from content
- Easy to apply 508-compliant styles
- Allows for lightweight content and presentation code

### Header Template

```
<html xmlns="http://www.w3.org/1999/xhtml">
<head>
  <title>Test codebook</title>
  <link href="codebook.css" rel="stylesheet" type="text/css" />
</head>
<body>
  <div id="content">
    <!-- End of Header -->
```

### CSS

```
body {
  font-family: Arial, Verdana, Sans-Serif;
  font-size: 8pt;
  margin: 0 auto; padding: 0;
  width: 400px;
}
hr { width: 100%; }
.variableInfo { margin-top: 10px; }
.variableInfo table { width: 100%; }
```

### Frequency Table Template

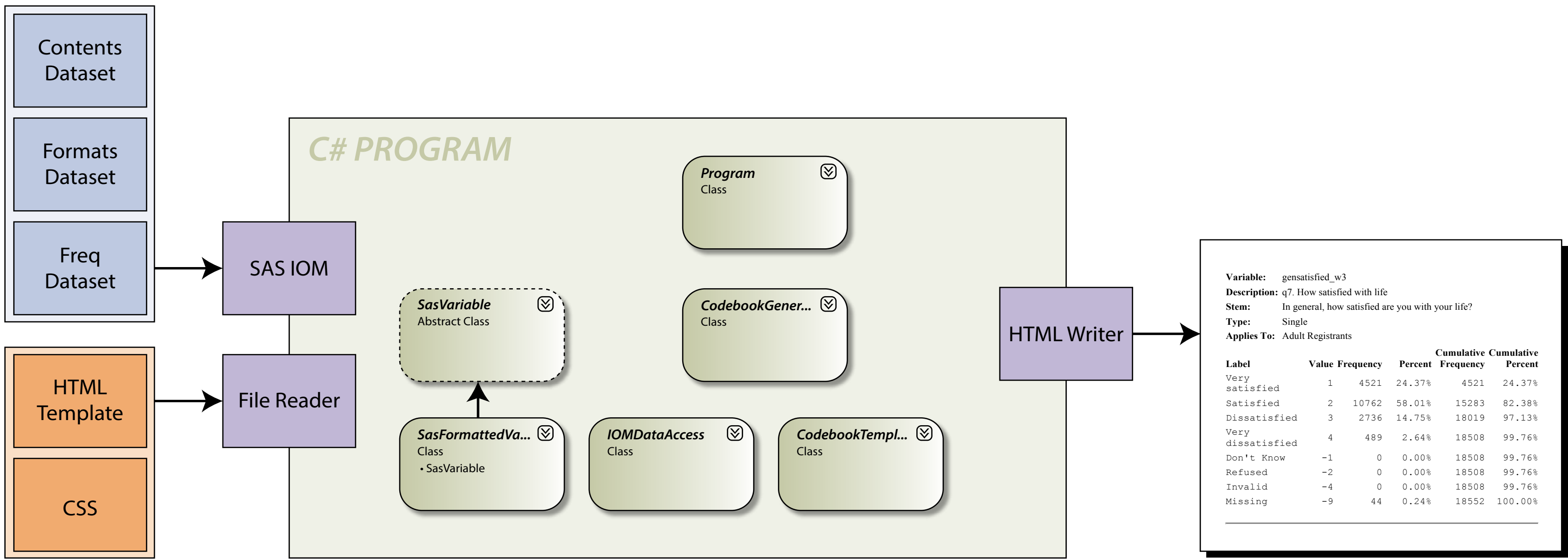
```
<div class="statistics">
  <table>
    <tr>
      <th class="leftJustify" width="35%"><span class="label">Label</span></th>
      <th class="rightJustify" width="5%"><span class="label">Value</span></th>
      <th class="rightJustify" width="20%"><span class="label">Frequency</span></th>
      <th class="rightJustify" width="20%"><span class="label">Percent</span></th>
      <th class="rightJustify" width="20%"><span class="label">Cumulative Frequency</span></th>
      <th class="rightJustify" width="20%"><span class="label">Cumulative Percent</span></th>
    </tr>
    {5}
  </table>
</div>
```

**Templates are easily modified with a simple text editor and can be customized for content**



## 5. Generating the Codebook

A .NET program, implemented in C#, is responsible for reading the codebook data, merging the data into a set of HTML templates, and producing the codebook file in HTML format. The program uses CSS styles to control the presentation of the codebook. The use of HTML templates and CSS stylesheets is an important design element of this program since they provide ability for codebook customization and many other benefits, as shown below. In addition, many modern programming constructs were used in the implementation, such as object-oriented programming, OLE DB, SQL, and SAS IOM.



SAS creates variable information, then C# applies the final style and template

## 6. Conclusion

A portion of a sample codebook is shown below. Some of the highlights are:

- Data Display
  - Double-column layout
  - Table layout for data
  - Heading labels in bold
  - Use of fixed-sized font to align frequency data
  - Horizontal break line between variables
- Usability
  - Easy to read
  - Conveys information clearly and distinctly
  - Reflects explicit research questions
  - Can be used as a summary to perform initial research exploration
  - Easily incorporated into word processing applications for reporting

<b>Variable:</b> ControlNo					
<b>Description:</b> Control number					
<b>Stem:</b> enrollee ID					
<b>Type:</b> Single					
<b>Applies To:</b> Adult Registrants					
</					

The final codebook result is easy to read and allows for multiple formatting options

## Contact Information

\*Presenting author: Helen Smith

For further information: Mai Nguyen Email: mnguyen@rti.org  
Phone: 919.541.8757 • Fax: 919.541.6178

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 trademarks of their respective companies.

