SAMHSA Web Guidelines
DMS-IT Web Technical Guidelines

Table of Contents

SAMHSA Web Guidelines
DMS-IT Web Technical Guidelines

Last Revised June 14, 2005

1. Approval and Update Procedures

1.1. Pre-Development Meeting

A meeting with the SAMHSA Webmaster, the Office of Communications (OC), the Government Project Officer (GPO) and their Technical Contractor Staff must be held prior to any Website development. This meeting is held for a general debriefing of the Website development and implementation procedures as well as expected deliverable standards to be followed. The SAMHSA Web Guidelines and the SAMHSA Checklist for New/Migrated Websites are provided to the GPO and their Technical Contractor Staff at this meeting. Once the Checklist has been completely filled out, it must be submitted to Webmaster@samhsa.hhs.gov. In addition, the official HHS/SAMHSA Clearance Form 524A must be submitted to OC before any Website Development begins.

Any technical specifications of the Website setup and design (including software packages to be used, programming languages to be used, client requirements, etc.) must be submitted to the SAMHSA Webmaster at this time.

In addition, functional requirements need to be discussed to determine if specific policies, procedures and regulations need to be followed.

NOTE: Any changes to the functional requirements discussed in this meeting, prior to deployment, must be communicated to the SAMHSA Webmaster. An email detailing the changes (i.e. new forms, new databases, personal information collection, product sales, new coding styles,) will need to be sent to Webmaster@samhsa.hhs.gov.

1.2. Website Hosting

Websites developed for SAMHSA are required to be housed on SAMHSA's servers, once development has been completed. Exceptions to this rule may be given, based on a projects scope and other factors, by the SAMHSA Webmaster.

1.3. Website Approval Process

Website design and layout (including the site map) must be approved by SAMHSA once a prototype of the Website has been completed. This prototype is typically placed on a SAMHSA Beta Server or on the Contractor's Development Server if hosted offsite.

In order to request a prototype approval, the Contractor will send an email to the Government Project Officer (GPO) and the SAMHSA Webmaster making them aware that the prototype is ready for review.

Once the Website development has been completed, the GPO must request a formal and final review of the Website. This review is conducted by SAMHSA's OC before the official launch.

***File formats:** When submitting files to SAMHSA DMS-IT staff for conversion to the Web for SAMHSA Hosted Sites, the only type of files that will be accepted are Microsoft Office files such as MS Word, MS Excel, MS PowerPoint, etc.

1.4. Website Post/Update Procedures - SAMHSA Hosted Sites

New content must be submitted to the Webmaster via email for posting to SAMHSA's Beta Server for review by Government Project Officer (GPO). DMS-IT staff will assume that the GPO has cleared this material with the OC if it contains any content listed under "Examples of Agency Information that Require Content Clearance" in the SAMHSA Web Guidelines on Page 2. Once the GPO has approved the new content, the SAMHSA Web Team will migrate the changes to the live Website. All emails to Webmaster must include:

- A zip file containing only the files that need to be posted to the site,
- A description of the content / db / being added to the site,
- A cc: sent to the GPO, and
- A cc: sent to the OC via Alvera.Stern@samhsa.hhs.gov.

Updated content must be submitted to the Webmaster via email for posting to SAMHSA's Beta Server for review by SAMHSA GPO. Once the GPO has approved the updates, the SAMHSA Web Team will migrate the changes to the live Website. All emails to Webmaster must include:

- A zip file containing only the files that need to be posted to the site,
- A description of the content /db/ code updates being made, and
- A cc: sent to the GPO.

All communications, including update procedures, are conducted with SAMHSA support staff via the "Webmaster@samhsa.hhs.gov" email account.

The SAMHSA Webmaster must receive notification by the GPO of any changes in Contractor who will be submitting content updates or updates will not be posted.

*** NOTE:** New content is defined as stated in the "SAMHSA Web Guidelines" as well as any additional web pages (files) that have been added to the site and were not previously present on that site**.**

1.5. Website Post/Update Procedures - Third Party Hosted Sites

New content* must be posted to the Contractor's Development Server for review and approval by the GPO BEFORE posting to the live public site. DMS-IT staff will assume that the GPO has cleared this material with the OC if it contains any content listed under "Examples of Agency Information that Require Content Clearance" in the SAMHSA Web Guidelines on Page 2. Once the GPO has approved the updates via written communication, the Contractor may migrate the changes to the live Website.

All communication regarding updates should be sent via email to the GPO and include:

- A description of the content /db/ code updates being made,
- The URL where the beta site can be viewed,
- A cc: sent to the SAMHSA Webmaster at Webmaster@samhsa.hhs.gov, and
- A cc: sent to the OC via Alvera.Stern@samhsa.hhs.gov for content clearance.

Updated content must be posted to the Contractor's Development Server for review and approval by the GPO BEFORE posting to the live public site. Once the GPO has approved the updates via written communication, the Contractor may migrate the changes to the live Website.

All communication regarding updates should be sent via email to the GPO and include:

- A description of the content /db/ code updates being made,
- The URL where the beta site can be viewed, and
- A cc: sent to the SAMHSA Webmaster at Webmaster@samhsa.hhs.gov.

*** NOTE:** New content is defined as stated in the "SAMHSA Web Guidelines" as well as any additional web pages (files) that have been added to the site and were not previously present on that site**.**

FOIA COMPLIANCE REQUIREMENT: FOIA regulations require SAMHSA to have backup copies of all public Websites. For this reason, the Contractor must submit to SAMHSA DMS-IT a complete and full backup of the Website every six (6) months. The backup should include all content, related code, databases, and graphical images used to build the site. Backups should be delivered to the SAMHSA Webmaster on CD.

2. Domain Names and Mail Box Requests

2.1. Domain Name Requests

SAMHSA follows the HHS IRM Policy for Domain Names (HHS-IRM-2000-0002). This policy is available at www.hhs.gov/read/irmpolicy/index.html.

2.2. Third-level Domain Names

Third-level domain names, such as <program name>.samhsa.gov or www.samhsa.gov/<program name> are acceptable URLs under this policy. Below outlines the procedure for obtaining a new third-level domain name.

- Third–level domain names can only be requested by the authorized Government Project Officer (GPO);
- The GPO must obtain prior approval for all domain names from the SAMHSA Office of Communications (OC); and
- Once approved, the PO must send an email to the SAMHSA Webmaster via Webmaster@samhsa.hhs.gov requesting that the approved domain name be set up.

2.3. Second-level Domain Names

Second-level domain names, or .gov addresses which omit "samhsa" such as <program name>.gov, require clearance by the Division of Management Systems-Information Technology (DMS-IT), Office of Communications (OC), and the Department of Health and Human Services (HHS). Justification for the waiver must be provided to DMS-IT, who will in turn request the waiver from HHS.

NOTE: New .org, .net, or .com domain names are not permitted for new production websites. In addition, **SAMHSA staff and related Contractors are not permitted to register .org, .net, .edu or .com names that directly mirror any of the SAMHSA third-level domain names.** For example, if a SAMHSA website name is csat.samhsa.gov, employees of SAMHSA and/or its contractors are not permitted to separately register the name csat.org, or csat1.org, csat.com, or any other variation of that same name. Previously registered site names such as health.org have been grandfathered into this policy.

2.4. Website Mail Box Requests

Website mail box requests must be submitted by the GPO to the IT Service Center and a government staff member must designated to monitor and administer the email box (e.g. Info@coce.samhsa.hhs.gov). The procedure for requesting an email box is outlined below:

- All email accounts must be requested by the Website GPO;
- The GPO must send an email to the IT Service Center (HHS/OS) and requesting that a Resource Email box be setup and include the government representative that will be responsible for administering the email box;
- A cc: of this email must be sent to the DMS-IT via Harvey.Karch@samhsa.hhs.gov .

- Once the email address (coce@samhsa.hhs.gov) is setup by the IT Service Center it can then be used on your Website.

3. Infrastructure Standards

3.1. Hardware Requirements

Any unique or specific hardware requirements beyond the norm for a website (e.g., proprietary hardware, large volumes of hard drive space, extra server(s)), require the prior approval from DMS-IT.

3.2. Software Requirements

The following is a list of software that SAMHSA currently supports for Website development. Prior approval from DMS-IT is required if software, other than what is listed below, is to be used for Website Development.

- **Web Server Operating Systems**: Windows 2000, IIS 5+
- **Database Technologies**: SQL Server 2000, Oracle 8+
- **Browser Technologies**: MS Internet Explorer (IE) 6+, Netscape Navigator 7+
- **Server Technologies**: ASP, ASP.NET, Oracle 9i Application Server, Cold Fusion 7 – (For existing systems only)

3.3. Prohibited / Restricted Web Technologies

Unless prior written approval is obtained from the SAMHSA Webmaster, Website and applications developed by SAMHSA contractors (on or offsite) are prohibited from using the following:

- Lotus Notes "Domino" Websites
- Cold Fusion Websites (Require Special Approval from DMS-IT)
- Microsoft FrontPage Extensions
- 'Forums' or 'Bulletin Board' areas on public SAMHSA websites (pages where users can directly upload content to SAMHSA web pages; if allowed, content posted must be held pending approval by a content expert)
- Real-time web chat and Instant Messenger
- List Servs
- RealAudio/Video

These restrictions apply to any technology that cannot be implemented using the software indicated in Section 3.2.

New Web Technologies are constantly emerging. It is not DMS-IT's intent to be restrictive about these technologies. Rather, our intent is to provide sensible solutions to the needs of each site. By the same token, we cannot possibly offer or support every web application or component on the market. If we do not currently offer a technology you would like to use, we may decide to do so based on joint needs, capabilities, and other factors. It is essential, however, that the Government Project Officer (GPO) ensures that the Websites meet DMS-IT requirements before development work begins.

3.4. SAMHSA Beta Server Access

SAMHSA's Beta Server are for internal staff use only, and can only be accessed from the SAMHSA Network, which is inside the SAMHSA firewall. If Contractors wish to view their site on SAMHSA's Beta Server, they must come into the SAMHSA building to do so.

External contractors, with the exception of the DMS-IT contract staff, are not authorized to perform work on any SAMHSA Server. On occasion, HTTP access may be granted to external Contractors for the purposes of viewing the Beta Site using a Web Browser. Such access is rare. Permission must be obtained via the SAMHSA Webmaster.

3.5. Administration Tools

IIS Web-based Administration tools are disabled on SAMHSA public web servers. Management is accomplished by the Webmaster or designated through the Microsoft Management Console (MMC).

In addition, the Cold Fusion (CF) Administrator will not be installed on any virtual sites that are available to the Public. CF Admin is available on its own virtual site which has been firewall restricted to only SAMHSA staff.

NOTE: ***URLSCAN** is a Microsoft tool that sits at IIS's "front door" and examines HTTP requests to Web Servers, allowing or denying them based on a rule set. The rule set denies most common hacker scripts from reaching the Web Server and consuming valuable IIS resources.

4. Website Development Standards

4.1. General Website Development Standards

Unless decided otherwise by the SAMHSA Webmaster (with input from the OC), new sites will be posted as a new virtual web, facilitating:

- ease of management / reporting,
- limited 'bloating' under www directories, and
- separation of processes for monitoring and troubleshooting.

New Virtual Webs are hosted using an IP address versus Host Headers, unless decided otherwise by the SAMHSA Webmaster. No duplication of files is allowed. Links must be used instead.

4.2. Browser Requirements

All Websites must be designed for 800x600 screen resolution and must function properly in both Netscape Navigator 7 (or greater) and MS Internet Explorer 6 (or greater).

4.3. HHS / SAMHSA Branding

All SAMHSA funded Websites must have the SAMHSA name on the site Logo and also in the footer of all Website pages. The footer should also have a link back to the SAMHSA Home Page at www.samhsa.gov.

Contractors who create draft logos must submit the logo to their GPO for approval. Logos must also be sent to the OC for review/approval. Requests for OC approval are normally submitted by the GPO.

4.4. Website Cookies

SAMHSA follows the HHS IRM Policy for Usage of Persistent Cookies (HHS-IRM-2000-0009). This policy is available at www.hhs.gov/read/irmpolicy/index.html and states:

"Persistent" web cookies shall not be used on HHS Websites, or by contractors when operating Websites on behalf of HHS agencies, unless the following conditions are met:

- The site gives clear and conspicuous notice;
- There is a compelling need to gather the data on the site;
- Appropriate and publicly disclosed privacy safeguards exist for handling any information derived from the cookies; and
- The HHS Secretary gives personal prior approval for the use.

"Persistent" web cookies are defined as web cookies that can track "the activities of users over time and across different Websites."

"**Session" web cookies** do not fall within the scope of this policy. Exempted cookies include those that retain information only during the session or for the purpose of completing a particular online transaction, without any capacity to track users over time and across different Websites. (Examples: for using shopping carts to purchase a number of items online or for filling out applications that require accessing multiple web pages.)

4.5. 508 Compliancy

All government funded Websites must comply with the requirements of Section 508 of the Rehabilitation Act (29 U.S.C. 794d). Section 508 requires that when Federal agencies develop, procure, maintain, or use electronic and information technology (EIT), Federal employees with disabilities have comparable access to and use of information and data as Federal employees who have no disabilities, unless an undue burden would be imposed on the agency. Section 508 also requires that individuals with disabilities, who are members of the public seeking information or services from a Federal agency, have comparable access to and use of information and data as the public without disabilities, unless an undue burden would be imposed on the agency. For more detailed information go to http://www.Section508.gov.

508-compliance is the responsibility of each Website's development staff. The official software used to check for 508 compliance by HHS and SAMHSA is WatchFire Bobby. The following must be adhered to:

- All sites developed must function properly in both Netscape Navigator 7 (or greater) and MS Internet Explorer 6 (or greater).
- If third-party browser plug-ins for Netscape or Internet Explorer are needed, prior testing and approval of the plug-ins by DMS-IT is required. Requests should be sent to the SAMHSA Webmaster at Webmaster@samhsa.hhs.gov.
- Linking to external Websites, which do not comply with Section 508 Accessibility regulations, is allowed provided that an Exit Disclaimer is used. (See Exit Disclaimer Section of this document for text to be used).
- All Website files must be coded to allow viewing from within the Web Browser without the use of additional programs or plug-ins. Examples of acceptable file format are: HTML, ASP, ASPX, DHTML, TXT, PDF, etc.
- If a PDF file is used on a Website, a text equivalent must also be provided.
- Reminder: All graphic files that directly relate to the context of a document must have a text equivalent available.

4.6. 208 Compliancy

All Websites developed must comply with the E-Government Act of 2002, Section 208. As of June 2005, SAMHSA translated all of their Human-Readable Privacy Policies into a standardized Machine-Readable format. During this process, SAMHSA was able to identify 1 master human-readable privacy policy (P3P) for all SAMHSA-hosted Websites. It is the responsibility of the Website Developers to make sure that the Website being developed complies with these policies.

In order to collect personal data the following needs to be taken into consideration:

- OMB Clearance is required for any form, survey or database designed to collect personal information from Website users.

- Section 208 Compliancy must be adhered to for all forms, surveys, and databases used for the purpose of collecting personal information. They also must be reviewed and approved by the SAMHSA Webmaster before going live.

- SAMHSAs Human-Readable Privacy Policies can be found at http://www.samhsa.gov/privacy.aspx.

4.7. Security and Restricted Access

Internet Security is a broad-based and complex topic. Therefore, this policy does not include significant guidelines on that subject. For any Internet Security issues, SAMHSA follows the HHS IT Service Center's Policy on Internet Security.

Restricted Access to a particular area on the SAMHSA Website is allowed with approval from the Office of Communications and the Division of Management Systems - Information Technology. User ID's and passwords for this secured area are created by the IT Service Center. A meeting with OC and DMS-IT will need to be conducted to work out all of the details of this secured area.

4.8. Exit Disclaimer

All links that point to a non-government / military sites (.ORG, .COM, .NET, .EDU, .TV etc.) must have an **Exit Disclaimer** that appears in the user Web browser before being redirected to the new web page. The text on the disclaimer message should say:

> "You are about to leave the SAMHSA website. SAMHSA provides links to other Internet sites as a service to its users, and is not responsible for the availability or content of these external sites. SAMHSA, its employees, and contractors do not endorse, warrant, or guarantee the products, services, or information described or offered at these other Internet sites. Any reference to a commercial product, process, or service is not an endorsement or recommendation by the SAMHSA, its employees, or contractors. For documents available from this server, the U.S. Government does not warrant or assume any legal liability or responsibility for the accuracy, completeness, or usefulness of any information, apparatus, product, or process disclosed."

NOTE: Linking to external Websites, which are (.GOV) or (.MIL), is allowed without a disclaimer.

4.9. Website Directory Structures (Onsite Hosting)

SAMHSA's Website Directory Structures are managed by using the following Directory Structure layouts:

- Sites hosted on SAMHSA servers are placed under D:\Webs\<projectname> (no spaces).

- All databases used by webs (where possible) are pulled from shared web directories and placed in D:\databases.

- All web log files will be pointed to D:\IISLogs. A text file will be placed in each subfolder, with the name of the corresponding web.

- User-friendly folder and filenames are encouraged because they ultimately become the URL structure.

4.10. Naming Conventions for Website Files

Web page URL's are a direct reflection of the Web site's file and folder names (ex. http://intranet.samhsa.gov/budget/fy2005.aspx). Having plain language folder names helps site visitors navigate the site, and relocate a previous page if they somehow get lost on the site, or have "jumped" to a location deep inside the web site from a search engine or other source.

The naming conventions used for the SAMHSA Website Files are as follows:

- No Spaces will be used in directory or filenames on public websites. Existing cases may remain at the discretion of the Webmaster.

- No more than one 'period' (.) may be used in a web directory or filename (for example, about.us.html is not allowed. aboutus.html is OK). Files with more than one period are denied by **URLSCAN*** for security reasons.

- No files with the following extensions are permitted on SAMHSA public Websites: (.exe, .com, .bat, .cmd, .ini, .log, .pol, .dat, .htw, .ida, .idq, .htr, .idc, .shtm, .shtml, .stm, .printer)

    1. **NOTE**: .shtm file includes should be renamed with an .html extension.
    2. All will be denied by **URLSCAN*** for security reasons.
    3. Executables as downloads may be made available in .zip format.
    4. Exceptions may be allowed in the future at the discretion of the Webmaster.

- None of the following URL sequences (i.e. sequence of characters typed into a browser address line) are permitted (all denied by **URLSCAN**):

  ..,
  ./

\
:
%
&

NOTE: '&' has been allowed due to existing cases (OAS), but future sites must adhere to naming conventions set forth.

4.11. Permissible Browser Requests

The following are permissible 'HTTP Methods' (browser requests):

- GET, HEAD, POST, DEBUG, TRACE

WEBDAV extended HTTP Methods such as the following are denied by **URLSCAN**:

- PROPFIND, PROPPATCH, MKCOL, DELETE, PUT, COPY, MOVE, LOCK, UNLOCK, OPTIONS, SEARCH

4.12. Avoid Duplication

To minimize duplication and improve the public's ability to locate accurate information across the array of government web sites, all SAMHSA funded web sites must link to existing government-wide portal or specialized sites when applicable, rather than re-creating these resources themselves.

Implementation Guidance:

- Before creating new information, the content managers must determine if that same – or similar – information already exists within their agency or on another federal web site.
- When a web site provides information or services for which there is a corresponding government-wide portal or specialized site, they must link to the government-wide portal or site from its pages on that topic.
- When a government-wide portal or specialized web site is available on a subject that the public would expect to find on an agency's site – but the agency does not provide that information – the agency must link to the government-wide portal or site in a logical and useful location.
- When content is the same or similar within agencies or across agencies, those agencies should consult with each other to find ways to share or coordinate content and to mitigate duplication.
- **IMAGE FOLDER:** All images and other multimedia files must be stored in the /IMAGES folder located off the root directory of the Website. Sub folders should be created topically (ex. FLAGS) so that content creators can easily find images when creating a page.
- **DATABASE FOLDER:** For all MS Access database used on the web site, the database file itself must reside in the /Database folder specified by the System Administrator.

4.13. Web Pages Must Link Back to the Home Page

To improve Web site usability, every federal web page must link back to its home page.

> **Implementation Guidance:** Many people do not recognize that an agency's logo links to the home page. If an agency uses only a graphical link, it must contain text indicating that it links to the home page.

4.14. Page Flow

Maintaining a consistent page flow is an important way to optimize a website or web-based application for search engines and optimize information dissemination for users.

Implementation:

Good page flow can be achieved by following some simple rules:

- **Page Headings:** First, use <H1><H2><H3><H4> (heading) tags to denote headings. The higher the number, the smaller the page heading. Much like Microsoft Word and PowerPoint, some spiders rip pages apart while indexing them and create their own table of contents for your document, and use heading tags to pick out major chapters and start/stop points.

- **Site Architecture:** Page flow also refers to the architecture of the site, and how other pages are linked within. If your navigation is straight forward and clean, chances are good that spiders will have an easy time indexing your site and documents, and all of the pages you wish the spider to follow will be followed.

4.15. Search Engine Usage

SAMHSA web sites are not required to use a search engine. However if a Search Engine is to be implemented as a feature for a particular web site, that site must point to the SAMHSA search engine located on www.samhsa.gov unless there is a compelling reason to build a separate customized search engine.

Customized search engines must obtain approval from the Government Project Officer and the SAMHSA Webmaster before development can begin.

4.16. Search Engine Optimization

SAMHSA's Internet sites must be optimized for crawling by standard Internet search engines such as Google, Yahoo and Open Source directories. SAMHSA's Internet sites must read and implement the three sub-sections below:

- Standard metadata implementation
- Using a robots.txt File
- DOCTYPE, Character Sets, and META Tag

4.17. Web Pages Must Use Standard Metadata

Metadata provides a standardized system to classify and label web resources. Meta tags improve search relevancy, let visitors know who created the information and when it was created, and allows information to be tracked and assembled by search engines. Metadata must be used on all second and third-tier web pages (one and two levels below the home page):

Implementation: At a minimum, each web page must include the following five metatags:

- Page Title,
- Creator (in most cases, the division name),
- Language,
- Publication Date,
- Subject and Keywords.

4.18. Using a Robots.txt File

All Internet Sites, and any other internal website that is to be included in the SAMHSA search engine, shall have a Robots.txt file stored in the top level (root) directory of the web site.

The Robots.txt file optimizes web application or web sites for search engine crawls. By following the instructions in this file, search engines can be instructed to ignore archived files or obsolete information on the website. Developers can also include development files and related directories as part of the robots file if the robots.txt file indicates those files and folder should be ignored.

Implementation: Here is an example of a basic robots.txt file:

Figure 1 – Example robots.txt file

```
1  # /robots.txt file for http://webcrawler.com/
2  # mail webmaster@webcrawler.com for constructive criticism
3
4  User-agent: webcrawler
5  Disallow:
6
7  User-agent: lycra
8  Disallow: /
9
10 User-agent: *
11 Disallow: /tmp
12 Disallow: /logs
```

4.19. How to Disallow Folders or Files From Being Crawled

The robots.txt file above was built specifically to inform robots and spiders that there are certain directories on this server that the webmaster doesn't want parsed or indexed.

- The first two lines, starting with '#', specify a comment

- The first paragraph specifies that the robot called 'WebCrawler' has nothing disallowed: it may go anywhere.

- The second paragraph indicates that the robot called 'lycra' has all relative URLs starting with '/' disallowed. Because all relative URL's on a server start with '/', this means the entire site is closed off.

- The third paragraph indicates that all other robots should not visit URLs starting with /tmp or /log. Note the '*' is a special token, meaning "any other User-agent"; you cannot use wildcard patterns or regular expressions in either User-agent or Disallow lines.

4.20. How to Disallow One Specific Web Page from Being Crawled

To prevent a spider from indexing or caching a certain page, add the following line of code to the <HEAD> of your document:

```
<META NAME="ROBOTS" CONTENT="NOINDEX">
```

Additionally, if you don't want the links on that page followed (and then indexed) by the spider, you should add the following META tag to the <HEAD> of your document:

```
<META NAME="ROBOTS" CONTENT="NOFOLLOW">
```

4.21. DOCTYPE, Character Sets, and META tags

Another important thing to remember when building applications and websites that are going to be indexed in search engines is that the HTML needs to be well-formed, and also needs to include some basic tags & attributes. The term "well-formed" means that the HTML is valid (according to w3 standards), and contains no errors. Here is an example of a perfect HTML 4.01 Transitional source (no content):

Figure 2 – Basic HTML layout

```
1  <!DOCTYPE HTML PUBLIC "-//W3C//DTD HTML 4.01 Transitional//EN">
2
3  <html lang="en">
4  <head>
5      <title>Page Title</title>
6      <meta http-equiv="Content-type" content="text/html; charset=UTF-8">
7      <meta http-equiv="Pragma" content="no-cache">
8      <meta name="description" content="">
9      <meta name="keywords" content="">
10 </head>
11
12 <body>
13
14
15
16 </body>
17 </html>
```

- On line 1 of the source code above, you'll see the DOCTYPE declaration. This is *extremely* important to include on **ALL** web pages and application pages. A DOCTYPE informs browsers and validators of what version of HTML (or XML, or XHTML) you are using, and must always appear as the very first line of code on the page.

- DOCTYPE declarations are essential to the proper rendering and functionality of web documents and applications in standards compliant browsers, such as IE 6, Safari, Firefox, Mozilla, and Opera.

- Using an incomplete or outdated DOCTYPE – or no DOCTYPE at all – throws browsers into what is commonly called "Quirks" mode, where the browser assumes you've written invalid or deprecated markup. In this setting, the browser will attempt to parse your page in backward-compatible fashion, rendering your styles as they might look in IE 4.0, reverting to a proprietary, browser-specific DOM. While some people may feel this is a good thing, it is actually a very bad thing. After all, one of your goals as a developer or designer is to make your content usable in as many formats as possible. If you build an application or website that only works in IE, you are cutting off the growing number of users who've switched to Firefox or Opera, or are using a different OS, such as Mac OS X or Linux.

- Without having the correct DOCTYPE declaration, it is very possible that this page will be ignored or incorrectly indexed by search engines, especially Google.

- On line 3 of the code above, you'll noticed that the standard <HTML> tag has the lang="en" attribute attached to it. What this does is to declare to the browser that this page is in English, and should therefore be parsed in:

    - Only English-language fonts, and
    - Should be translated into English by text-readers and other browser add-ons.

- On line 5 is the <TITLE> tag. It is ESSENTIAL for search engine optimization that EVERY PAGE has a UNIQUE title. The <TITLE> tag is used by search engines on the results page. It's used to hyperlink to the page, and is the first thing the user sees when they parse the results. If your <TITLE> tag isn't descriptive, or is the same for every page, your search results will be confusing.

- Another essential tag that should be on ALL pages is the <META> content-type declaration, which tells the browser which character set it should use to parse your pages. This is found on line 6 of our example above.

- We use the UTF-8 character set because it is the most universally recognized, and works very well with legacy systems. This exact tag should appear, as is, on all web pages.

- On line 7 of the example, you'll see another <META> tag called the "Pragma" tag. This tag, while not essential, serves a very important function. It tells browsers and spiders NOT to cache the page in their history, so that every time they visit the site, they'll get the newest content. Again, this tag is not essential, but it's very helpful in preventing people from caching data.

- Lines 8 and 9 are your basic META Description and Keyword tags. Much like the <TITLE> tag, each page needs to have a UNIQUE description and keywords, so that it can be indexed properly. The description should always be a quick summary of what that page shows, as well as keywords that are related to the content within the page.

For instance, if the page you are working on is for CMHS and relates to the latest grant information for psychiatric institutions, here's the description and keywords you would use:

```
<META name="description" content="2005 CMHS grant information for mental health institutions">
<META name="keywords" content="samhsa, 2005, grant, mental, health, psychiatric, information">
```

- The Meta elements "description" and "keyword" are part of the Dublin Core Metadata Element Set (version 1.1 as of 2/1/2005). The Dublin Core (as it is referred to in shorthand) are a set of Meta elements used to uniquely identify a document, it's content, it's creator(s), it's format, and many other elements. There are a total of 15 elements in the Dublin Core. Those elements (and their definitions) are:

Element Name	Description/Purpose
Title	The name given to the document
Creator	Author or person who maintains the document
Subject or Keywords	Topic's and relevant keywords related to the document
Description	Short summary of the contents of the document
Publisher	Name of the person who is responsible for making the document publicly viewable.
Contributor	A person or people who contributed to the creation of or the content within the document
Date	Typically used for the date of last update to the document or its contents
Resource Type	General category, genre, or aggregation level of the content
Format	MIME type of the document (for web – text/html commonly)
Resource Identifier	Typically indicates the URL (Universal Resource Locator), URI (Uniform Resource Identifier), or other unique string of characters/integers used to identify the document. ISBN is also an example
Source	Similar to Resource Identifier. Typically used to identify the original source of the information contained within the document.
Language	Language which the document is published in. It is suggested that RFC 3066 (http://www.ietf.org/rfc/rfc3066.txt) should be used to code this element.
Relation	Reference information to a related resource
Coverage	Extent or scope of the content of the document, such as a time period, location, jurisdiction, or other similar identity
Rights Management	Information about whether the document is covered by copyright or other intellectual property data

- Typically, only a few of the Dublin Core elements are used for web pages and applications. The most common are keyword, description, title, date, and publisher.

Usage of these core elements should be determined on an organizational or departmental level.