

## DETAILED METHODOLOGY



### The Web Site Evaluation Process

In 2009, CMF, with the assistance of our research partners at Harvard Kennedy School, Northeastern University, University of California–Riverside, and the Ohio State University, conducted an extensive evaluation of all congressional Web sites in the 111th Congress. The evaluations included 439 House Member Web sites,<sup>1</sup> 99 Senate Member Web sites,<sup>2</sup> 68 House & Senate Committee Web sites (majority and minority), and 14 House & Senate Leadership Web sites.

Our evaluations focused solely on the official Web sites of Congress and the experience and expectations of a typical end-user. We did not review or assess the following, which were outside the scope of this project:

- **Politics, policies, or positions.** We looked how effectively Web sites were providing content and information, and not at the individual merit of any positions.
- **Intranet, institutional, support, or Member or staff organization Web sites.**
- **Adherence to accessibility standards and the rules of their respective chambers.** We strongly advise that all congressional sites follow these guidelines, but we did not have the capabilities to evaluate adherence to them.
- **Back-end design or technical components of the Web site,** including whether or not an office uses, or has used, a vendor for design, maintenance, and/or hosting.
- **Congress's online presence independent of their official Web sites.** For instance, we did not subscribe to and evaluate e-newsletters, or search for Twitter accounts if they were not linked to from the official site. While we believe Members should have an online presence beyond their official Web site, it was outside the scope of this project.

---

<sup>1</sup> Includes 433 Representatives (there were two vacancies at the time of our evaluations), 5 delegates, and 1 resident commissioner.

<sup>2</sup> There was one vacancy in the Senate at the time of our evaluations.



## Developing the Criteria

To determine how to best judge the quality of a congressional Web site, CMF conducted research from a variety of sources, including:

- Focus groups with citizens.
- Interviews with Members of Congress and management, legislative, administrative, support, and technical staff from the House and Senate.
- Surveys of reporters and advocates.
- Research on industry best practices and usability.

Through this research, we determined that successful Web sites follow five core principles, which we then used to develop specific criteria. The five core principles are:

1. **Know your audience(s).**
2. **Provide timely and targeted content that meets their needs.**
3. **Make the site easy to use.**
4. **Foster interaction both on and offline.**
5. **Add value through innovation.**

A visitor's experience on a Web site is largely based on qualitative factors such as the quality and tone of the information, the usability, navigability, look and feel of the site, and the degree to which the information meets the user's needs. As a result, a purely objective measurement-based evaluation does not begin to assess actual user experience. To comprehensively evaluate the user experience we utilized a methodology that combines quantitative and qualitative measurements of the sites. Our evaluation methodology was designed to be as fair and objective as possible without being based solely on objectively-measured factors.

Through the five core principles and our extensive research, we identified specific features or qualities that, taken together characterize high-quality Web sites in the normative sense. That is, Web sites that are helpful and informative from the perspective of all users. Each question, called a "code," tapped into one of the attributes identified as important for a Web site. A form used by all evaluators provided a detailed description of the attribute that the code was intended to measure and guidance on how to respond to the question. Some of the codes are qualitative — asking how well the site did something — such as the National Issues code: "To what extent does the site provide information about major national issues..." and the evaluator is to respond on a 0 (low) to 5 (high) scale. Other codes are quantitative — asking whether or not a site had something — such as the presence or absence of a THOMAS search box, which is measured as either present (one) or absent (zero).

The criteria were developed for the following reports. Prior to each evaluation year, the criteria are updated and refined to reflect current and evolving technology and practice standards.

1. *Congress Online: Assessing and Improving Capitol Hill Web Sites* (published January 2002)
2. *Congress Online 2003: Turning the Corner on the Information Age* (published March 2003)
3. *2006 Gold Mouse Report: Recognizing the Best Web Sites on Capitol Hill* (published February 2007)
4. *2007 Gold Mouse Report: Lessons from the Best Web Sites on Capitol Hill* (published January 2008)
5. *111th Congress Gold Mouse Project* (published April 2010)


Each category of sites — House and Senate Member, committee, and leadership — has a different evaluation form based on the different roles, goals, and audiences of each category of office. Every site within the three categories was evaluated extensively. In 2009, Member Web sites were judged on 93 criteria in the following broad categories:

1. Timeliness
2. Information on Issues
3. Usability
4. Constituent Services and Casework
5. Accountability
6. Legislative Process
7. Floor Proceedings
8. District/State Information
9. Communication Technology
10. Media Communication

Because of the unique and divergent nature of each committee and leadership site, the evaluation of these sites necessarily relied less on a large number of quantitative codes and more on qualitative codes specific to each site's purpose and goals. The 61 committee criteria and 49 leadership criteria fell into most of these categories as well, but were adjusted to reflect their unique roles (for example, they were obviously not judged on District/State Information or Constituent Services and Casework).

## Ensuring Fairness and Accuracy

For the congressional Web site evaluations, every evaluator went through several rounds of training to ensure that he or she accessed the site and judged each criterion — especially the qualitative ones — in the same way, within a reasonable margin of error. The training included everything from assuring that all evaluators used the same browser (Internet Explorer version 7) to assuring that all evaluators judged the difference between a "4" rating and a "5" rating in the same way.



After extensive training, evaluators then received a randomly generated list of Web sites by our research partners and evaluated them in that order. The evaluations of the Web sites of all Members of the House and Senate were done by six evaluators between June 29 and August 13, 2009. The evaluations of the Web sites of all committee and leadership sites of the House and Senate were done by a single evaluator between September and December 2009.

## Scoring the Web Sites

After all 620 congressional Web sites were evaluated, the resulting data was analyzed with the assistance of our research partners. Though detailed information on the specific way each category of site was scored follows, they all were scored through the same basic steps:

- **Criteria were weighted** according to importance and factored into a formula that resulted in an overall numerical score for the site.
- **Scores were sorted** within each category — Senate Member, House Member, Committee, and Leadership — standardized, and translated into letter grades.
- **A Mouse Award was given** to the 135 sites that scored an A- or higher. There was no pre-set number of awards. The top-ranked site in each category was recognized with the first-ever Platinum Mouse Award.

### MEMBER SITES

While it is relatively straightforward to identify and code for the attributes that make for a good quality Web site, determining how to aggregate these data to measure the overall quality of a Web site is more difficult. One possible and relatively easy way would be to add all of the coded variables together and create a simple percentage of the number of codes on which a Web site received a high score. This “additive index” approach is not desirable, though, since it gives equal weight to each code. For example, in this approach, whether or not the site had information on receiving federal grants — which is somewhat peripheral to the intrinsic quality of a legislative Web site — would be given equal weight to whether or not the Member offered rationales for key votes or extensively discussed national issues.

Instead, we used statistical factor analysis to assign weights to the codes and then constructed the numeric overall quality score by taking the weighted sum of the coded variables.

Factor analysis is one method social scientists use when there are many measured indicators of a trait that itself cannot be measured directly, such as a Web site’s intrinsic “quality.” A trait such as “Web site quality” that cannot be measured directly is called a “latent” trait. Factor analysis is a statistical method that takes many coded variables, called “indicators” or “indicator variables,” and finds the latent trait that

is common to all of the indicators.<sup>3</sup> Here, the measured quality indicators are the coded variables, and the latent trait of the Web site is its underlying quality. Factor analysis is especially appropriate in this context, since it accommodates the fact — indeed assumes — that many of the indicators are measured subjectively or with error. With the presence of a vast quantity of indicators, subjectivity and errors tend to cancel out.<sup>4</sup>

We carried out the statistical analysis in several steps:

1. First, we used factor analysis and the coded data to construct a series of 13 attribute scales, where each attribute scale measured the extent to which the Web site possessed desirable attributes such as the amount of issue content, the Web site's usability, the quality of the information to constituents, and the quality of the Web site's technology. For this analysis, we used 59 codes, and between three and six coded indicator variables to measure each attribute scale.
2. Second, we combined these attribute scales using a separate factor analysis model to form two quality scales, one measuring the quality of the information on the site, and the other measuring the degree to which the site communicates to constituents.
3. Third, we took the average of the two quality scales to construct an overall quality score.

Figure 1 lists all of the coded variables we used in the analysis, and the composition of each of the attribute and quality scales. The original coded variables are indicated with a square bullet point (▪) (for example, the code for "Readability"). The attribute scales are indicated with a circular bullet point (•) (for example, "The quality of issue content on the site"). The two quality scales are labeled as such (for example, "Information Quality Scale").

---

<sup>3</sup> Factor analysis uses the empirical correlations among indicator variables to estimate the degree to which the original data can be described by one or more underlying dimensions. These underlying dimensions to the data are the latent trait or traits of interest, and the observed, coded data are used to measure the unobserved latent traits. Since the variables in this study are either dichotomous (0,1) or ordered categorical (0 to 5), we use a polychoric correlation matrix in the factor analysis. We use the principle factors method, and for each attribute scale we select variables that load well onto only a single dimension for that scale. We retrieve factor score coefficients using the regression method, which post-multiplies the vector of factor loadings by the inverse of the estimated correlation matrix. These scoring coefficients are the weights we use to construct each attribute scale. For ease of presentation, we standardize the scoring coefficients as proportions within each scale so they add up to one.

<sup>4</sup> While factor analysis statistically accommodates data that are measured with error, it is best for the data to have as little error as possible, or in other words, high "reliability." We conducted a reliability study across the coders by asking each coder to evaluate 10 common Web sites. Six coders were involved in the study. Coders evaluated all Web sites in a randomized order, and so did not know which 10 Web sites were the common Web sites. To assess reliability across these common evaluations, we quantified a coding "error" as choosing a value that differed from the modal response from all coders. So for example, if coder B coded a Web site as having an attribute, and so entered a one, while the remaining eight coders entered a zero, coder B in this instance would be identified as having made a coding error. We then simply took the percentage of errors across all codes that each coder recorded for all common Web sites. The error rates overall were very low, ranging from as low as 9.6% to as high as 13.8%. This indicates overall a very high level of inter-coder agreement. The strong reliability of these data reflects the extensive training each coder received. The factor analytic methods that we use identify and correct for any errors in the coding, so that coding errors do not affect the grade a Web site receives.

**FIGURE 1. FACTOR WEIGHTS FOR THE CODED INDICATORS**

**INFORMATION QUALITY SCALE**

- **The quality of all issue content on the site (0.23)**
  - The quality of information regarding national issues†
  - The quality of the Member’s priority issues
  - The quality of state and local issues
  - The extent of the Member’s rationales given for votes cast†
- **The amount of content on current issues (0.33)**
  - The quality of information on national issues from the 111th Congress†
  - The quality of Member’s priority issues from the 111th Congress
  - The quality of state and local issues from the 111th Congress
  - The extent of rationales for votes from the 111th Congress
  - The Member’s specific accomplishments from the 111th Congress
- **The ambiguity of the issue content (-0.11)**
  - The discussion of national issues is too generic to tell if current
  - The discussion of the Member’s priority issues is too generic to tell if current
  - The discussion of state and local issues is too generic to tell if current
- **The Web site’s usability (0.24)**
  - The overall look and feel of the Web site†
  - Ease of navigation†
  - Readability†
  - Organization†
  - Timeliness of information and updates†
- **The timeliness of the information (0.31)**
  - National issues are from the 111th Congress
  - Member’s priority issues are from the 111th Congress
  - State and local issues are from the 111th Congress
  - Rationales for votes are from the 111th Congress
  - The Member’s specific accomplishments are from the 111th Congress
  - Video from the 111th Congress
  - Audio from the 111th Congress
  - Timeliness of information and updates†

**COMMUNICATION QUALITY SCALE**

- **The extent to which the Web site promotes accountability to constituents (0.17)**
  - Provides the Member’s voting record
  - Explains how to read and use roll call votes
  - Lists the Member’s sponsored and co-sponsored legislation



- **Information on the legislative process (0.16)**
  - Explains how a bill becomes a law
  - Includes a FAQ section about Congress
  - Provides information about what a Member does
  - Includes a student or kid's page
- **Information on floor proceedings (0.14)**
  - Lists the chamber schedule
  - Provides information about the current floor proceedings
  - Links directly to the *Congressional Record*
- **Resources for constituent casework (0.13)**
  - Answers FAQ regarding constituent casework†
  - Provides guidance on how constituents can initiate casework
  - Includes a casework privacy release form
  - Links to relevant government agencies
  - Links to USA.gov
- **Information of interest specifically to constituents (0.07)**
  - Congressional internships
  - Admission to the military academies
  - Capitol tour information
  - Flag requests
  - Information about federal grants
- **Information about the district or state (0.13)**
  - Lists interesting and important features of the district/state
  - District/state demographic statistics
  - Includes map of the district/state
  - Information about local or district resources
- **Media communication (0.10)**
  - Press releases are organized by date
  - Video
  - Audio
  - Archives of the Member's columns or op-ed pieces
- **Communication technology (0.10)**
  - Blog
  - RSS feed
  - Online poll or survey
  - Facebook
  - Social Bookmarking
  - Twitter

Variables marked with a † are measured on a 0 to 5 scale; all others are measured on a 0 to 1 scale.

Each scale has an associated weight listed in parentheses, where the weight is estimated in a factor analysis model, and then rescaled so that the weights add up to 1. With this rescaling, the weights indicate the proportions each item contributes to the associated scale.

To construct the attribute factor scores, we multiplied the value of each coded variable by its associated weight, and then added these numbers together. Most of the codes are measured on a (0/1) scale, meaning that the variable is a one if the Web site has the feature and a zero if the Web site does not have the feature. Because of this, the attribute scales for the most part are simply the sum of the weights corresponding to the items found on the Web site.

To create the quality scales for “information” and “communication,” we multiplied the Web site’s factor score for each of the relevant attribute scales (that is, the attribute scales that fall under each of the two headings) by its associated weight and added these together. Finally, to create a total score, we took the average of the Web site’s score on the two quality scales.

Both of the quality scales measure the quality of a Web site but each in different ways: one measures issue content and the other measures the clarity of communication to constituents. While our analysis showed that these are distinct dimensions of quality, they are not unrelated. The two quality scales were highly correlated (with a Pearson correlation of 0.49,  $p < 0.001$ ).

Through our research and experience we identified four key dimensions most essential to a successful Web site which we believe should be given priority beyond the results generated in the factor models. These four dimensions are issue content, constituent casework, timeliness, and usability. Web sites that scored above average on all four of these dimensions were given extra credit in the form of a 10% increase above the overall score generated by the factor analysis.

The final step was to convert the numeric scores to letter grades. To do this, we first sorted all scores within each category, so that the scores were arranged from the highest House Member score to the lowest House Member score, and then the highest Senate Member score to the lowest Senate Member score.<sup>5</sup> We then computed a within-chamber ranking by dividing each Web site’s score by the baseline score for its category. In this final ranking, sites that did better than the best site in their category last year received a score above 100%. All of the other scores ranked below that as the percent of the baseline score. We then maintained past cutoffs between scores to assign letter grades.

## COMMITTEE AND LEADERSHIP SITES

Committee and leadership sites, like Member sites, were evaluated using most of the qualitative and quantitative measures described above, with the criteria honed to their particular purpose. To judge all sites with the most accurate and fair evaluation possible, we divided committee and leadership offices into “classes.” These classes were designed to take into account the differing needs of the audiences that

<sup>5</sup> We conducted an overtime reliability analysis to ensure comparability of the scores between the 2007 and the 2009 data. Twenty-two of the 30 codes tested had an error-rate less than 16%, most with no error (the remaining 8 items had an error rate of 33% or more, however). Overall, these results demonstrate good overtime reliability.




the committees each serve: some committees primarily serve more technical and professional audiences, while others primarily serve the general public. Thus, one class of committees must be held to higher standards for providing information geared toward a Hill audience and lower standards for providing information accessible to the general public, while a second class of committees should be held to higher standards for information geared to the general public and lower standards for the information useful for a Hill audience. The use of these classes enabled us to use the same criteria in the evaluation of all of the sites while still ensuring a robust and accurate scoring of each individual site.

## Committees

The classes of committees were based on the primary audience(s) they serve. All committees serve the same basic audiences — the general public, the press, experts on the issues in their jurisdiction, and congressional staffs. Committees differ, however, in the degree to which particular audiences demand access to their work. As a result, we separated committees into the following five classes:

1. **Standing committees with significant professional audiences.** These committees have jurisdictions that are fairly specialized and/or technical and whose audiences will include a significant percentage of people who have some knowledge of, and expertise in, the subject matter. *These committees were held to lower standards for providing educational and other information geared toward the general public, but higher standards for providing technical documents and publications geared toward an expert audience.* Examples of committees in this class include the House Committee on Appropriations and the Joint Economic Committee.
2. **Standing committees with significant non-professional audiences.** These committees have jurisdictions that impact a fairly sizeable population of citizens with a personal, rather than an academic or professional, interest in the subject matter. *These committees were held to higher standards for providing information and services accessible to the general public, but lower standards for providing technical information for an expert audience.* Examples of committees in this class include the Senate Committee on Veterans' Affairs and the House Small Business Committee.
3. **Standing committees with significant professional and non-professional audiences.** These committees have jurisdictions that make their work of interest to a range of expert and non-expert audiences. Most of them have high media profiles, which keeps their work in the public eye. *These committees were held to high standards for having information and services accessible both to the public and to experts.* Examples of committees in this class include the House Committee on Energy and Commerce and the Senate Committee on Health, Education, Labor, and Pensions.
4. **Standing committees with primarily congressional audiences.** These committees have jurisdictions that are largely internal to Congress. *These committees were held to high standards for providing information and services for congressional staffs and lower standards for providing information and services for external audiences.* It is important to note, however, that some of the functions of these committees are best served by internal congressional intranets, which were not included as part of our evaluation. Committees in this class include the Senate Committee on Rules and Administration and the Committee on House Administration.

- 
5. **Standing committee minority offices with primarily partisan audiences.** Minority offices of committees do not control the bulk of the information that most visitors will be seeking on committee Web sites and provide a counterpoint to the majority office. *These committees were held to high standards for providing information and services for their Republican staff, supporters, and the press.* Examples of committees in this class include the Senate Committee on the Budget (Minority) and House Committee on Foreign Affairs (Minority).

## Leadership Offices

The classes for leadership offices were based on their leadership roles. Their different roles dictate the audiences they are trying to reach and the content and services they should be providing. All leadership offices have to provide basic educational information about what they do; information about the Leader's and the party's message; key issues, information and services for same-party congressional staff and Members; and information in support of their leadership roles. The degree of attention they should devote to each of these areas should vary, however, based on their particular responsibilities. To allow for this in our evaluations, we divided leadership into two classes:

1. **Leadership offices with both congressional and public audiences.** These leadership offices hold the highest leadership positions of each party in each chamber. Their roles are to represent the party, as a whole, and to provide general party leadership on and off Capitol Hill. They are the spokespersons for their parties, and their Web sites will be of interest to citizens as well as to party activists and congressional staffs. As a result, *they were held to high standards for providing authoritative party information for both citizens and congressional staff.* Note that the Majority and Minority Leaders in the Senate do not have separate leadership Web sites. Examples of leadership sites in this class include the House Republican Leader and the House Majority Leader.
2. **Leadership offices with primarily congressional audiences.** These leadership offices conduct the organizational business of the parties. Though their content and services — and their Web sites — are of interest to select citizens and activists off Capitol Hill, their primary role is to organize the party on Capitol Hill. As a result, *they were held to high standards for providing content for congressional staffs and lower standards for providing information and services that are accessible to citizens.* An example of leadership sites in this class is the House Republican Policy Committee.

The evaluation criteria were weighted based on their overall importance, and the scores for the audiences were given extra weight based on the class of the site. As was done with the Member grades, the scoring was standardized to ensure comparability across the years. We computed a ranking by dividing each Web site's 2009 score by the highest 2007 score within the respective category (committee or leadership). In this final ranking, the highest scoring Web site in 2007 for each category was given a 100%, and all of the other scores were ranked as the percent of that score. The cutoffs between scores for committee and leadership sites, established in the *2006 Gold Mouse Report*, were used to assign letter grades.