



**Web Metrics and Evaluation:  
Current State of Implementation  
Among the CENDI Agencies  
PHASE 1**

*Submitted by*  
**CENDI Metrics and Evaluation Task Group**

*Sponsored by*  
**CENDI User Education Working Group**

*Prepared by*  
**Gail Hodge**  
**Information International Associates, Inc.**  
**Oak Ridge, Tennessee**



*August 2000*

# TABLE OF CONTENTS

---

---

Executive Summary .....	1
1.0 Introduction.....	3
2.0 Summary Matrix.....	3
3.0 Agency by Agency Review .....	5
3.1 Department of Energy, Office of Scientific and Technical Information ....	5
3.2 Defense Technical Information Center .....	5
3.3 National Agricultural Library .....	6
3.4 National Library of Education.....	6
3.5 National Library of Medicine .....	7
3.6 National Technical Information Service .....	9
3.7 US Geological Survey, Biological Resources Division.....	9
4.0 Findings by Topic.....	9
4.1 Metrics and Definitions.....	9
4.2 Evaluation of User Satisfaction .....	10
4.3 Internet Connectivity/Performance Metrics .....	11
5.0 Related Activities by Other Groups .....	12
6.0 Recommendations .....	13
<i>Appendix A Web Metrics and Evaluation Survey.....</i>	<i>i</i>
<i>Appendix B Interpretation of Results from NAL's Web Statistics .....</i>	<i>ii</i>
<i>Appendix C Definitions from ED's Implementation of WebTrends.....</i>	<i>vi</i>
<u>Tables/Matrices</u>	
Summary CENDI Metrics and Evaluation Matrix .....	4
Detailed Matrix on Metrics and Definitions .....	10
User Satisfaction Matrix .....	10
Detailed Matrix on Internet Connectivity/Performance Metrics .....	11

## **CENDI METRICS AND EVALUATION TASK GROUP**

Terry Hendrix (DTIC)

Gail Hodge (CENDI Secretariat)

Ed Lehmann (NTIS)

Linda Tague (NLE)

Mark Thompson (USDA/NAL)

Chalmers Wilson, Kay Sutton (DOE/OSTI)

Dr. Fred Wood (NIH/NLM)

Lisa Zolly (USGS/BRD)

CENDI is an interagency cooperative organization composed of the scientific and technical information (STI) managers from the Departments of Agriculture, Commerce, Energy, Education, Defense, the Environmental Protection Agency, Health and Human Services, Interior, and the National Aeronautics and Space Administration (NASA).

CENDI's mission is to help improve the productivity of Federal science- and technology-based programs through the development and management of effective scientific and technical information support systems. In fulfilling its mission, CENDI member agencies play an important role in helping to strengthen U.S. competitiveness and address science- and technology-based national priorities.

## EXECUTIVE SUMMARY

Phase I of this CENDI activity, Web Metrics and Evaluation, is complete. The task group has collected information on relevant activities from the participating CENDI agencies and established a new baseline. Since the last CENDI review of this topic (1998), CENDI agencies have further intensified their use of web log analysis software for generating usage data. All agencies make use of such software. Several CENDI agencies have extended their web evaluation program to include usability studies, focus groups, expert reviews, and/or online surveys. The concept of web evaluation is clearly maturing, and the use of web metrics is growing. General interest in web metrics and evaluation is increasing, both in the government and among a wide range of commercial, academic, research, and various nonprofit, user, and Internet provider organizations. With regard to web performance and Internet connectivity, the Defense Technical Information Center (DTIC) and the National Library of Medicine (NLM) now have active programs to monitor performance based on both internal and external data. Web/Internet performance evaluation was in its infancy two years ago. Today, bandwidth and peak congestion issues are well recognized, although solutions remain elusive. Many agencies are still not actively engaged in addressing “end-to-end” Internet connectivity as an important dimension of overall web site performance.

This changing web landscape presents CENDI with several needs and opportunities. First, while web metrics data are widely collected by CENDI agencies, there is, as yet, no common framework of agreed upon metrics. Nor is there a central location for access to such data or a mechanism for sharing knowledge about web metrics. CENDI agencies seem ready to move forward now on developing a web metrics framework. Second, while CENDI agencies are engaging in more diverse web evaluation activities, there is no continuing mechanism for sharing information and experiences. This is especially the case with regard to online surveys of web users, a relatively new and underdeveloped area that has every indication of expanding rapidly in the near future. CENDI agencies seem ready to learn about and engage more sophisticated “second generation” online survey options. Third, the Internet connectivity aspect of web performance is receiving less attention compared to other aspects of web evaluation. Most CENDI agencies do not have an end-to-end connectivity perspective or program, and what activities there are tend to be network, backbone, or ISP-centric rather than user (end-to-end) oriented.

We suggest the following priorities for possible consideration in Phase II of this activity:

- Develop a common web metrics framework--a CENDI web watch. This would address a standard core of web metrics/definitions, best metrics practices, and related metrics research or other supportive research.
- Develop a state-of-the-art online survey knowledge base. This would cover survey instrument development, Office of Management and Budget (OMB) approval process, and survey implementation issues.
- Develop a tutorial overview on Internet/web performance monitoring, including options for collaborative Internet performance monitoring--perhaps a CENDI Internet watch.
- Plan and convene one or more workshops on Items 1, 2, and/or 3.

- Decide on a CENDI strategy for engaging STI/information policy issues from a web-centric perspective (e.g., privacy, Freedom of Information Act [FOIA], Government Performance and Results Act [GPRA], permanent public access, WebGov). Define as Phase III of this task group, or establish a separate group.

## 1.0 Introduction

Increasingly, the CENDI agencies have been asked to quantify the uses of, satisfaction with, and impact of their products and services. Much of the impetus is coming from increased use of the Web where it is more difficult to know the users, particularly if the service does not require user registration for access. In addition, metrics and evaluation are important for product development and innovation, resource allocation, capacity planning, and to respond to Government Performance and Review Act (GPRA) requirements.

Based on the importance of metrics and evaluation, the CENDI members proposed that this be an area of effort for CENDI in FY00. This effort is an update to the previous Metrics and Evaluation Task Group report (CENDI/98-1) published in 1998.

In November 1999, a task group was formed under the auspices of the User Education Working Group. This group has membership from 6 of the 10 CENDI agencies, including NTIS, NLM, DTIC, USGS, NLE, and NAL.

The group developed a survey instrument based initially on the survey used in the 1998 study (see Appendix A). The survey was completed by the members of the task group and sent to the CENDI Principals and Alternates of agencies without direct representation.. Of these, DOE/OSTI provided contributions to this report.

This report is Phase 1 of a multi-part project. The report's purpose is to provide a baseline for the current state of metrics and evaluation at each agency that contributed. It compares and contrasts the software, metrics, and reports that are currently in use. At the end of the report, there is a brief discussion of issues that have been identified during this analysis. This will provide input for the follow-on work in Phase 2, which will attempt to identify the gaps in what is being measured/tracked and to make recommendations for what might be done to fill these gaps.

## 2.0 Summary Matrix

This report focuses on three aspects of web metrics – usage, evaluation, and performance. Usage metrics measure various aspects of the frequency and types of uses of agency web sites. Evaluation metrics focus on the customer utility, usability, and satisfaction. Performance metrics measure the speed and efficiency of providing the information, whether displaying a page, downloading a file, or performing a transaction. In each case, the information collected in this study documents both the current activities in these areas and future plans.

A summary of the information collected from the participating agencies is presented below. More detail is provided in the following sections by agency and by topic.

## SUMMARY CENDI METRICS AND EVALUATION MATRIX

Question/Agency	DOE/OSTI	DTIC	NAL	NLE	NLM	NTIS	USGS
<b>Number of Web Sites Monitored</b>	40 organizational homepages; over 1000 sites	> 90; manage but do not report on all	1 and subdirectories some of which are monitored individually	> 200; manage and report information	6	17	Focused on NBII Site only
<b>Software for Usage Analysis</b>	Access	Access	HTTP-Analyze	WebTrends Enterprise Edition	HTTP-Analyze, Analog, WebTrends, and custom in-house	NetTracker, Statbot, and WebTrends	WebTrends
<b>Future Software for Usage Analysis</b>	Analog				FunnelWebPro		
<b>Current Evaluation Activities</b>		Developed draft performance measures	Informal observations of trends and frequently visited pages	Syracuse Univ. study and online customer survey	Various units have carried out: focus groups, usability studies, expert review, external review, online bounceback surveys	Web statistics are tracked by the web site's content managers	Online web survey
<b>Future Evaluation Activities</b>	Commercial software to better position web products	Consultant will help better define and implement metrics	None	Next redesign of site will involve more up-front usability testing and customer focus groups	Web Evaluation Work Group is considering web usage monitoring services (e.g., with a large panel of Internet users who have agreed to have their web usage monitored), and vendor and academic surveys of users of health information on the web.	None	Developing standards for the biological informatics community that will give minimal information across all sites. A partnership manual is also being developed.
<b>Connectivity / Performance</b>	Reports from ISP	Compare site performance to others; usage statistics based on router utilities	Reports from ISP	RedAlert service used to ping sites for downtime	Commercial and in-house software to monitor performance; monitors selected links between NLM and various locations; partnerships to test and evaluate general Internet performance.		
<b>Future Performance Activities</b>	None		None	None	The NLM Web Evaluation Work Group will develop a plan for the next stage of performance evaluation research. Likely to include ongoing operational capabilities as well as testing experiments		

### 3.0 Agency by Agency Review

#### 3.1 *Department of Energy, Office of Scientific and Technical Information (DOE OSTI)*

The Department of Energy, through its various Program Offices, maintains more than 40 organizational home pages. From these sites, more than 1000 additional sites are maintained with new sites added as needed. Within the Department, various tools and collection methodologies are currently being utilized to collect, analyze, and assess information. As the Department progresses towards standardization of its information infrastructure, a baseline for tools, metrics and evaluation criteria is expected to evolve for acceptance and adoption.

DOE OSTI currently uses Access software for Web log analysis, but is moving to Analog. The current software collects the number of hits and the number of pages requested. These are broken down by domain, by time frame (week, month or year), and unique hosts.

Statistical software is used to check link popularity. DOE believes that search engine readiness needs more attention. This includes the impact of meta tags, titles, etc. It is difficult to position products to show up in the top 10. They are looking into commercial software to better position their Web products with search engines (e.g., Web Position Gold). Effective exit data (how do people exit your site) are also needed.

In terms of connectivity and performance, OSTI relies on its ISP (ESNet) for traffic information

#### 3.2 *Defense Technical Information Center (DTIC)*

DTIC manages over 90 web sites. The majority of these are “owned” by other DoD groups. In these cases, captured statistical data is used differently by DTIC and the owner. Content/access data is provided to the owner of the page; data transfer/server usage data is used by DTIC to manage server availability. The reports are available at <http://www.dtic.mil/usage/>.

DTIC uses Access Watch to collect this data, which collects unique hosts, both internal and external; the file requests by type; errors in completing user requests; the megabytes of information transferred; and the number of html pages per hour and day that they are demanded. Because DTIC manages sites for many other DoD groups and programs, they have agreed to provide specific access information to the site owners. Information is collected for all web pages.

DTIC has been interested in Performance Measures for some time. They have hired Dr. Tim Sprehe as a consultant and, together, have drafted *Performance Measures for Federal Agency Websites* in Nov. 1999. However, this is a conceptual document. In order to take it to the operational level, DTIC along with other federal agencies, are working together with Dr. Sprehe to develop web performance measures to be used across all of DoD. In addition, two reports are available that evaluate usage of DefenseLink (*DefenseLink: Heuristic Evaluation* (1997) and *DefenseLink Usability Study* (1997)).

DTIC’s evaluation of connectivity and performance uses the Keynote service ([www.keynote.com](http://www.keynote.com)) to compare DTIC’s site performance to those of 40 important business sites. Keynote calculates the average response time experienced by web users across the U.S. during

the hours of 6:00 am to noon, Pacific Time, Monday through Friday, from over 60 measurement locations. The CISCO router utility, Multi Router Traffic Grapher, also provides information about network performance. Daily, weekly, and monthly graphs are provided that show bandwidth utilization.

### **3.3 *National Agricultural Library (NAL)***

The NAL home page and the Schoolmeals web sites are monitored. Major subdirectories, such as AGRICOLA and the Information Centers, are monitored individually. NAL has written scripts that provide log data for the various NAL units.

NAL uses HTTP-Analyze. The metrics collected include hits, the number of files downloaded, the pageviews, sessions, and the number of kilobytes sent. The metrics are defined in Appendix B, Interpretation Results. Statistics are available to users within the <nal.usda.gov> domain. The statistics are presented both as raw numbers and graphically. Bar and line graphs are provided to show trends over time. In addition, the major statistics identified above are plotted in such a way that they can be analyzed together. Full statistics are also provided for each month. This includes response codes such as 404 (not found) and 403 (forbidden).

The average load by day is used as a baseline to compare the previous week. This is done for hits, files downloaded, and pages viewed. Raw numbers are provided for the top seven days of the period being reported. The average hits per hour are also provided in a bar graph.

NAL is observing trends and noting some of the most frequently visited pages. One information center reported an apparent drop in e-mail and written reference requests after posting a new segment on the web, which serves as an indirect way to show the benefits of these pages. Although there are no current plans for future activities in this area, some managers have expressed an interest in further web analysis of this type.

NAL is starting to see partial saturation of the T1 line and is hoping to upgrade. Performance information on bandwidth utilization is made available via periodic reports from NAL's ISP and is plotted by day. The mean, 50<sup>th</sup> percentile and 95<sup>th</sup> percentile are plotted. There are no specific plans to collect more detailed information.

### **3.4 *National Library of Education (NLE)***

NLE manages and reports on the main Department of Education (ED) web site and those of most ED offices and programs. Until recently, NLE managed the <ed.gov> web site for the Department of Education. NLE has provided reports on the main ED web site and those of most ED offices and programs. ED also monitors the usage of its search engine, Ultraseek, and the cross-site index of more than 200 ED-funded sites and more than 150 education-related, federal sites (part of the Federal Resources for Educational Excellence). The National Center for Education Statistics, the FAFSA on the web, which handles interactive student aid applications, the ED grants system, and the contracts and grants information, are also monitored. The Lotus Domino server, which houses several applications, is monitored separately.

ED uses the WebTrends Enterprise Edition to analyze web usage logs. WebTrends reports for many of the major ED web servers are available at <http://www.ed.gov/internal/webstats/>. In

addition to the overall report for the main site, <www.ed.gov>, ED produces subset reports for each principal office, and for specialized services such as RealMedia, web discussion forums, and SSL applications.

The metrics include the number of hits, the number of page views, visitor sessions, and the number of unique visitors. These terms are defined in Appendix C. Monthly Internet status reports give management overview of web activity, most used home page categories, most viewed pages, trends over time, busiest day-of-week/time-of-day, etc. The report also features a rotating focus on browsers and platforms used, referrals from Internet portals and search engines, and analyses of domains, time-of-day, day-of-week, etc. The March 2000 report is available in PDF form at <http://www.ed.gov/internal/rscd0003.pdf>.

ED has done significant work in the area of web site evaluation. In September 1998, the School of Information Studies at Syracuse University was commissioned to study selected ED web sites from four perspectives -- management assessment, policy analysis, web log and transaction analysis, and usability testing. The January 1999 report, "Evaluation of Selected Websites at the U.S. Department of Education: Increasing Access to Web-based Resources", is available from <http://www.ed.gov/internal/webeval/>. Since November 1996, more than 2,450 responses have been captured via the ongoing online customer survey (<http://www.ed.gov/Survey/cust.html>). An analysis of the results from September 1999 is available at <http://www.ed.gov/Survey/memo1999/>.

ED expects that the next redesign of the ED site during the second half of calendar year 2000 will involve more up-front usability testing and customer focus groups.

ED analyzes web site availability by reporting the amount of downtime recorded in the logs of the RedAlert service. This service pings several ED web sites and services (ED home page, search engine, databases, discussion forums, etc.) every 15 minutes, 24 hours a day, seven days a week.

### **3.5 National Library of Medicine (NLM)**

NLM uses HTTP Analyze, Analog, WebTrends, and custom in-house software to collect usage statistics. In the near future, FunnelWebPro will replace HTTP Analyze and Analog.

NLM home page ([www.nlm.nih.gov](http://www.nlm.nih.gov)); MedlinePlus ([medlineplus.gov](http://medlineplus.gov)); PubMed ([www.ncbi.nlm.nih.gov/PubMed](http://www.ncbi.nlm.nih.gov/PubMed)); Internet Grateful Med ([igm.nlm.nih.gov](http://igm.nlm.nih.gov)); Specialized Information Services ([sis.nlm.nih.gov](http://sis.nlm.nih.gov)); and Clinical Trials Database ([clinicaltrials.gov](http://clinicaltrials.gov)) are monitored. Data are generated for each of the monitored web sites. At present, NLM does not have an overall web monitoring framework or consistent reporting format. However, this is being developed by the NLM Web Evaluation Work Group, which is reviewing web metrics and definitions.

Data on the following metrics are or will be collected by log analysis (monitoring) software for one or more NLM web sites: number of pages downloaded (page downloads); requests for pages downloaded; searches (of specific searchable databases); unique visitors (per unit of time, e.g., day); sessions (defined time that a single user is logged on); total hits (but can be misleading due to images, headers, forms--overstates the number of users); frequency of use or visit (e.g., first

use, repeat use, x times per period of time--but limited to users with fixed IP addresses); length of use (time online per visit--also limited to users with fixed IP addresses); referring URLs (the immediately preceding link used to access a given web site); most frequent web pages requested (web pages on a site ranked in order of number and percentage of requests); search engines used (e.g., Yahoo, Excite, HotBot) to access a web site; topics searched; web browsers used (e.g., Netscape, Explorer) when accessing a web site; operating systems used (e.g., Windows 98, 95) when accessing a web site; user navigation paths within a site (e.g., link by link URL pathways for a specific user session); user domain name with country domain indicated (subject to a significant error factor due to foreign users using .com or .net); user domain name with organization domain indicated (e.g., edu, gov--subject to a significant error factor due to heavy use of com or .net).

Various NLM units have carried out: focus groups (MedlinePlus, PubMed); usability studies (MedlinePlus); expert review (MedlinePlus); external review (SIS); and online user feedback surveys (PubMed). Additional evaluations of these types are planned for the future. NLM has requested and received OMB blanket approval for customer satisfaction surveys of its web site users. Each individual survey will still require OMB approval, but on an expedited basis.

In addition, the NLM Web Evaluation Work Group is considering external web usage monitoring services (e.g., with a large panel of Internet users who have agreed to have their web usage monitored), and vendor and academic surveys of users of health information on the web. Some vendors offer online focus group or chat room services, as well as random sampling surveys of the client's web users. NLM hopes to be able to move ahead with some of these new activities in the near future.

Keynote Systems ( [www.keynote.com](http://www.keynote.com) ) is used to monitor the performance of selected NLM web sites. Keynote measures average download times from user emulation servers at about 60 U.S. and international Internet points-of-presence, and compares these download times with similar measurements for a group of well-known commercial web sites. Data are used for general performance monitoring, not as a defined performance benchmark or metric. In addition, NLM uses a variety of commercial, shareware, and custom software for its own monitoring of selected Internet links between NLM and various U.S. and international locations. Over the last couple of years, NLM has partnered with various U.S. and international academic and commercial organizations to test and evaluate Internet performance.

An earlier Internet study was published by F.B. Wood, V.H. Cid, and E.R. Siegel, "Evaluating Internet End-to-End Performance: Overview of Test Methodology and Results," *Journal of the American Medical Informatics Association*, Vol. 5, November/December 1998, pp. 528-545. A manuscript on results of recent NLM performance testing of high bandwidth pathways has been submitted for consideration by a leading engineering journal. An overview of NLM's Internet performance research is being prepared for inclusion in a manuscript to be submitted to a leading scientific journal.

Internet performance evaluation is within the purview of the NLM Web Evaluation Work Group. The Work Group will develop a proposed plan for the next stage of NLM's Internet performance evaluation research. This may include suggestions for an ongoing operational capability as well as the next round of testing experiments. The primary original contribution of NLM's work is the emphasis on end-to-end Internet performance, from the viewpoint of the end user. Most

Internet performance work to date in the commercial and academic research communities has focused on backbone or ISP or high-bandwidth “cloud” performance rather than end user-to-end user.

### **3.6 *National Technical Information Service (NTIS)***

NTIS runs thirty web sites, six of which are actual NTIS sites. The remainder of the sites are managed by NTIS for other agencies. NTIS uses NetTracker, a service that is hooked onto another server. Statistics are available from the NetTracker site. Statbot and WebTrends are also used internally. The web content managers for sites run by NTIS use the usage statistics gathered by these programs. For the NTIS main site, the NetTracker data is analyzed quarterly for trends in usage and to review the most popular pages. The performance of NTIS’ web sites is monitored by using KeyNote Systems at [www.keynote.com](http://www.keynote.com).

### **3.7 *US Geological Survey (USGS)***

While the USGS has many web sites, the input to this report focuses on the web site for the National Biological Information Infrastructure. This site is unusual because it is essentially a gateway site to other web sites. Web sites may be cataloged by the NBII staff at USGS or by other contributors. A product called Tag-Gen allows the staff and contributors to apply metatags to web sites.

Because of the distributed nature of the NBII contents, it is difficult to determine metrics or evaluation criteria. Each partner and, sometimes, each linked site has its own way of determining usage, customer satisfaction and performance. The NBII is working on a handbook for partners that will address metrics and evaluation. Minimal standards are being developed for biological informatics sites.

The NBII currently has a web-based user survey available. However, it has proven very difficult to get responses to this survey. It will likely be taken down and replaced with something that will, hopefully, elicit better response.

## **4.0 Findings by Topic**

Three main aspects of web metrics discussed in this report are web usage metrics and their definitions; evaluation of user satisfaction; and Internet connectivity and web site performance.

### **4.1 *Metrics and Definitions***

All agencies polled collect statistics regarding web usage. In most cases, the statistics are gathered based on commercial software. Some agencies have also added customized software to connect the statistics to certain systems or to provide customized reports or graphs. The detailed information about metrics and definitions is provided in the table below.

### DETAILED MATRIX ON METRICS AND DEFINITIONS

Agency	Metrics
<b>DOE/OSTI</b>	# of hits; # of page demands; broken down by domain; timeframe (week, month and year); and unique hosts
<b>DTIC</b>	Unique hosts; number from DTIC and number external; file requests by type; errors in completing user requests; megabytes of information transferred; html pages per hour and html pages each day (avg. estimate)
<b>NAL</b>	Hits; files downloaded; page demands; sessions; Kbytes transferred (see Appendix A for detailed definitions)
<b>NLE</b>	Hits; page views; sessions; unique visitors (see Appendix B for detailed definitions)
<b>NLM</b>	Number of pages downloaded (page downloads); requests for pages downloaded; searches (of specific searchable databases); unique visitors (per unit of time, e.g., day); sessions (defined time that a single user is logged on); total hits (but can be misleading due to images, headers, forms--overstates the number of users); frequency of use or visit (e.g., first use, repeat use, x times per period of time--but limited to users with fixed IP addresses); length of use (time online per visit--also limited to users with fixed IP addresses); referring URLs (the immediately preceding link used to access a given web site); most frequent web pages requested (web pages on a site ranked in order of number and percentage of requests); search engines used (e.g., Yahoo, Excite, HotBot) to access a web site; topics searched; web browsers used (e.g., Netscape, Explorer) when accessing a web site; operating systems used (e.g., Windows 98, 95) when accessing a web site; user navigation paths within a site (e.g., link by link URL pathways for a specific user session); user domain name with country domain indicated (subject to a significant error factor due to foreign users using .com or .net); user domain name with organization domain indicated (e.g., edu, gov--subject to a significant error factor due to heavy use of com or .net).
<b>NTIS</b>	Hits; files downloaded; page demands; sessions; number of visits (by time frame); usage summary and details by directory; usage detail by page
<b>USGS</b>	

The most common statistics gathered include number of pages requested, megabytes of files downloaded, and visitors by IP address groups. There is generally an attempt to separate internal use from external use. Generally, the information is gathered on hourly, daily, and weekly timeframes. Trends are often presented by year.

#### 4.2 Evaluation of User Satisfaction

Four of the seven agencies perform user satisfaction evaluations. Some have been done by focus groups, some through usability testing, some with surveys, etc.

#### USER SATISFACTION MATRIX

Agency	Evaluations
<b>DOE/OSTI</b>	None
<b>DTIC</b>	Small studies conducted several years ago. Primarily geared to web design and usability. Draft performance measures and hiring a consultant. : Recent Survey conducted on DTIC-owned and managed web page, Secure STINET.
<b>NAL</b>	Nothing formal other than observing trends and noting some of the most frequently visited pages.

<b>Agency</b>	<b>Evaluations</b>
<b>NLE</b>	Syracuse Univ. study evaluated 4 perspectives: management assessment, policy analysis, web log and transaction analysis, and usability testing.
<b>NLM</b>	Various NLM units have carried out focus groups, usability studies, expert review, external review, online surveys.
<b>NTIS</b>	Nothing formal other than observing trends and noting some of the most frequently visited pages.
<b>USGS</b>	The NBII has had a web survey online but it isn't working well. It is going to be redone.

Based on anecdotal evidence from the agency representatives, the most useful evaluation technique appears to be the more formalized surveys such as focus groups, because it is difficult to get people to complete online web surveys. However, a combination of techniques is likely to be the best approach.

Efforts are underway to evaluate how to migrate the various surveying techniques and methods into the Internet environment. Of particular interest is how to provide an online survey so that it is completed – where should it be placed, is redundant access from various places on the site of value, how should the users be sampled, etc.

#### ***4.3 Internet Connectivity/Performance Metrics***

Less than half of the agencies included in the study were involved in connectivity and performance metrics. Those who indicated some involvement were primarily using metrics provided by their ISP. Some were also using statistics provided by their router utility programs. Only DTIC, NLE and NLM had specific programs in place to routinely analyze connectivity and performance. All three used outside services – RedAlert or Keynote for this purpose. DTIC and NLM also use a variety of other metrics and software.

#### **DETAILED MATRIX ON INTERNET CONNECTIVITY/PERFORMANCE METRICS**

<b>Agency</b>	<b>Connectivity/Performance Metrics</b>
<b>DOE/OSTI</b>	Statistics from ISP
<b>DTIC</b>	Bandwidth utilization. Daily, weekly and monthly graphs; Collected at specific intervals. Compare site performance to others using Keynote; usage statistics based on Cisco router utility, "Multi Router Traffic Grapher".
<b>NAL</b>	Statistics from ISP
<b>NLE</b>	Analyzes web site availability by reporting the amount of downtime recorded in the logs of the RedAlert service, which pings several ED web sites and services every 15 minutes 7X24
<b>NLM</b>	Uses Keynote for external monitoring of web download time as a function of time of day as a primary metric. Uses a variety of other metrics and test software, including primarily: web download times and throughput; bulk transfer capacity (throughput or bandwidth); traceroute (number, location, and sequencing of hops); and ping (round trip time and packet loss). Has experimented with various commercial testing software, and recently extended the testing program to include both the commodity Internet and high bandwidth Internet pathways (e.g., vBNS).

Agency	Connectivity/Performance Metrics
NTIS	NTIS uses Keynote Systems to monitor performance. Monitoring is done on a regular basis.
USGS	None

## 5.0 Related Activities by Other Groups

Federal Webmasters' Forum has done nothing specific in the area of metrics, but is interested in what CENDI discovers ([www.it.policy.gsa.gov/mke/fedwebm/fedwebm.htm](http://www.it.policy.gsa.gov/mke/fedwebm/fedwebm.htm)). However, the Forum did have a presentation from the Government Accounting Office (GAO) on the need for new metrics in a transaction-based environment.

The W3C has a Web Characterization Working Group. The effort includes development of web characterization terminology and definitions (this work was part of OCLC's contribution to the W3C group). They will be working on the actual metrics in the near future. The Working Group also maintains a repository that has interesting presentations, white papers, and links to related web resources. Virginia Tech maintains the repository. These resources are linked from the W3C site at [www.w3c.org](http://www.w3c.org).

The Metrics Group of D-Lib Magazine ([www.dlib.org/metrics/public](http://www.dlib.org/metrics/public)) is also very active. The group is currently working on various metrics scenarios for the different functions performed within the web search environment. The most active sub-group is that on retrieval headed by Carl Lagoze at Cornell University. Many of the ideas that are being discussed within this subgroup may find their way into the recently awarded DL-2 project at Cornell.

The SIGMETRICS group within the American Society for Information Science ([www.asis.org](http://www.asis.org)) maintains a listserv that is an open list. There is also an archive of messages. The owner, Gretchen Whitney from University of Tennessee, has provided via the listserv many pointers to web sites and documents that address issues related to metrics. While the majority of them are of a standard bibliometric nature, there are some related to web usage.

OCLC's Office of Research has several metrics-related initiatives. One project measures the scope of the web. The Web Characterization Project gives statistics about the number of sites, the number with publicly available content, the number indexed, etc. Included in this project is a list of Metric Properties ([www.oclc.org/oclc/research/projects/webstats/currmetrics.htm](http://www.oclc.org/oclc/research/projects/webstats/currmetrics.htm)) that specifies the data collection unit and its scope in time and space. These definitions may be useful to this group's effort.

From the commercial side, there are several white papers and sites of interest from vendors. Of special note is the white paper by Keynote on "The Economic Impacts of Unacceptable Web Site Download Speed" ([http://www.keynote.com/downloads/download\\_main.html](http://www.keynote.com/downloads/download_main.html)). PC Data has some results available without cost from its site (<http://www.pcd.com>). Headcount.com, while similar to several other sites, is unique in that there are monthly "ask the experts" chat sessions where specific questions can be asked. It is just a matter of watching the list to see if the current expert is in your area of interest.

While the involvement in metrics by OMB, GAO, the Federal Webmasters' Forum, or the CIO Council appears to be minimal, indications are that the interest in web metrics on the part of these administrative organizations within the government will increase. The emphasis on government performance, customer satisfaction, efficiency and effectiveness, as well as increased use of the web as the basis for interaction between the government and its constituents is likely to increase the interest in related web metrics and evaluation.

## 6.0 Recommendations

Phase I of this CENDI activity has shown that the baseline has moved significantly since the 1998 CENDI review of web metrics. General interest in web metrics and evaluation is increasing, both in the government and among a wide range of commercial, academic, research, and various nonprofit, user, and Internet provider organizations. Activities in metrics and evaluation now include analysis of web logs, the incorporation of traditional customer satisfaction methodologies, such as focus groups and surveys, and web performance and Internet connectivity.

Since the last CENDI review of this topic (1998), CENDI agencies have further intensified their use of web log analysis software for generating usage data. All agencies make use of such software. A variety of software, both commercial and internally developed, is used. More metrics are being captured and their sophistication has increased. However, differences in the definitions of the metrics remain based on differences in the software and in the individual implementations by the agencies.

Several CENDI agencies have extended their web evaluation program to include usability studies, focus groups, expert reviews, and/or online surveys. The concept of web evaluation is clearly maturing. However, questions remain about how to take these traditional methodologies and techniques into an Internet environment.

With regard to web performance and Internet connectivity, DTIC and NLM now have active programs to monitor performance based on both internal and external data. Web/Internet performance evaluation was in its infancy two years ago. Today, bandwidth and peak congestion issues are well recognized, although solutions remain elusive. Most CENDI agencies do not yet have an end-to-end connectivity perspective or program, and what activities there are tend to be network, backbone, or ISP-centric rather than user (end-to-end) oriented.

This changing web metrics environment presents CENDI with several needs and opportunities. First, while web metrics data are widely collected by CENDI agencies, there is no common framework of agreed upon metrics. However, it appears that CENDI agencies could benefit from developing such a web metrics framework. Second, while CENDI agencies are engaging in more diverse web evaluation activities, there is no continuing mechanism for sharing information and experiences. This is especially the case with regard to online surveys of web users, a relatively new and underdeveloped area that seems destined to expand rapidly in the near future. CENDI agency representatives are interested in learning about and engaging in more sophisticated "second generation" online survey options. Third, the Internet connectivity aspect of web performance is receiving less attention compared to other aspects of web evaluation; most CENDI agencies are still not actively addressing "end-to-end" Internet connectivity as an important dimension of overall web site performance.

Based on these observations, the task group suggests the following priorities for possible inclusion in Phase II of this activity:

- Develop a common web metrics framework--a CENDI web watch. This would address a standard core of web metrics/definitions, best metrics practices, and related metrics.
- Develop a state-of-the-art online survey knowledge base. This would cover survey instrument development, the OMB approval process, and survey implementation issues.
- Develop a tutorial overview on Internet/web performance monitoring, including options for collaborative Internet performance monitoring--perhaps a CENDI Internet watch.
- Plan and convene one or more workshops on Items 1, 2, and/or 3.
- Decide on a CENDI strategy for engaging STI/information policy issues (e.g., privacy, FOIA, GPRA, permanent public access, WebGov) from a web-centric perspective. Define as Phase III activities for this task group, or establish a separate group.

## Appendix A

### Web Metrics and Evaluation Survey

#### Survey Questions

Please provide information on the following for your CENDI agency:

1. web-based usage monitoring software or approaches currently in use;
2. specific web usage metrics and definitions in use;
3. which web sites are being monitored;
4. copies (or URLs) of any recent, illustrative reports or data analyses on web usage;
5. web site evaluation activities (this includes web site design and web site customer satisfaction) currently in use, eg., focus group, expert review, usability lab, on-line bounceback survey, on-line chat room, off-line survey, commercial usage monitoring service, commercial on-line survey;
6. copies of any recent, noteworthy reports or data analyses on completed web evaluation activities;
7. future agency activities or initiatives re web evaluation;
8. any thoughts or findings on what is working best so far re web usage and web site evaluation and what issues need attention;
9. Internet connectivity and performance evaluation software or services currently in use (this includes both in-house and commercial);
10. specific Internet performance metrics and tests in use;
11. copies (or URLs) of any recent, noteworthy reports or data analyses on Internet performance;
12. future agency activities on Internet performance evaluation and thoughts on what is working best to date and issues that need attention.

## Appendix B

### Interpretation of Results from NAL's Web Statistics

#### Interpretation of Results

The statistics report contains among others the following information:

- the number of hits, 304's, files, pageviews, sessions, data sent (in KB)
- the amount of data requested, transferred, and saved by cache (in KB)
- the number of unique URLs, sites, and sessions per month
- the number of all response codes other than 200 (OK)
- the average hits per weekday and for last week
- the maximum/average hits per day and per hour
- the number of hits, files, 304's, sites, data sent by day
- the top 5 days, 24 hours, 5 minutes and 5 seconds of the summary period
- the top 30 most commonly accessed URLs (hits, 304's, data sent)
- the 10 least frequently accessed URLs (hits, 304's, data sent)
- the top 30 client domains accessing your server most often
- the top 30 browser types
- the top 30 referrer hosts
- the overview/detailed list of all files requested
- the overview/detailed list of all sites by domain and reverse domain
- the overview/detailed list of all browser types
- the overview/detailed list of all referrer URLs

The following section describes the meaning of all those numbers in the summary report which are not self-explaining:

**Hits** (color key: green) A hit is any response from the server on behalf of a request sent from a browser. This includes any response from the server, not only text files or documents. If, for example, a HTML page has two images embedded, the server generates three hits if this page is requested: one hit for the HTML page itself and two hits for the two inline images.

#### Files

(color key: blue) If the user requests a document and the server successfully sends back a file for this request, this is counted as a Code 200 (OK) response. Any such response is counted for as a file. Again, "file" here means any kind of a file.

**Code 304 (Not Modified)**

(color key: yellow) A Code 304 (Not Modified) response is generated by the server if a document hasn't been updated since the last time it was requested by the user and therefore there was no need to actually send the files for this document. This happens if the browser (or a caching proxy server between the browser and your web server) still has an up-to-date copy of the page in its local storage (cache) and therefore can display the page without requesting the actual content. This technique is used to reduce network traffic, but it also causes an inaccuracy in the statistics reports regarding the number of visitors, because the browser or proxy usually sends only one such a conditional request per user session if it still holds an up-to-date copy of the file. However, the ratio between "files" and "304's" reflects the efficiency of overall caching mechanisms for at least those hits which made its way to the server.

**Pageviews**

(color key: magenta) Pageviews are all files which either have a text file suffix (.html, .text) or which are directory index files. This number allows to estimate the number of "real" documents transmitted by your server. If defined correctly, the analyzer rates text files (documents) as pageviews. Those pageviews do not include images, CGI scripts, Java applets or any other HTML objects except all files ending with one of the pre-defined pageview suffixes, such as .html or .text. See also the Pageview directive.

**Other responses**

There are much more responses than only Code 200 (OK) and Code 304 (Not Modified) responses, especially in the coming standard, the HTTP 1.1 protocol specification. For example, the server could generate a Code 302 (Redirected) response if a page has moved, a Code 401 (Unauthorized Request) response if access to the document is denied or a Code 404 (Not Found) response if the requested page does not exist on this server. See the HTML specification at <http://www.w3.org/> for information about all valid responses from a web server. Note that http-analyze does recognize HTTP/1.1 responses according to RFC2068.

**KBytes transferred**

(color key: orange) This is the amount of data sent during the whole summary period as reported by the server. Note that some servers do log the size of a document instead of the actual number of bytes transferred. While in most cases this is the same, if a user interrupts the transmission by pressing the browser's stop button before the page has been received completely, some servers (for example all Netscape web servers) do not log the amount of data

transferred but the amount of data which would have been transferred if the user would have completely loaded the page.

**KBytes requested**

This is the amount of data requested during the whole summary period. http-analyze computes this number by summing up the values of KBytes transferred and KBytes saved by cache (see below).

**KBytes saved by cache**

The amount of data saved by various caching mechanisms such as in proxy servers or in browsers. This value is computed by multiplying the number of Code 304 (Not Modified) requests per file with the size of the corresponding file. Note: Because http-analyze can determine the size of a file only if the file has been requested at least once in the same summary period, the values for KBytes saved by cache and KBytes requested are just approximations of the real values.

**Unique URLs**

Unique URLs are the number of all different, valid URLs requested in a given summary period. This shows you the number of all different files requested at least once in the corresponding summary period.

**Unique sites**

This is the sum of all unique hosts accessing the server during a given time-window . The time-window is hardwired to the length of the current month. This means that if a host accesses your server very often, it gets counted only once during the whole month. Only the sum of the unique hosts per month is listed in the statistics report.

**Sessions**

(color key: red) Similar to unique sites, this is the number of unique hosts accessing the server during a given time-window. This time-window is one day by default for backward compatibility, but it can be changed with the option -u or the Session directive in the configuration file. For example, if the time-window is two hours, all accesses from a certain host in less than 2 hours after the first access from this host are lumped together into one session. All following accesses more than 2 hours apart from the first access will be counted as a new session. This way you may get an estimated number of how many sessions are started on different sites to access your server.

**Unresolved**

If you have disabled domain name lookups in your web server to decrease response times of your server or if the host isn't

configured in the Domain Name System (DNS) for whatever reason, http-analyze cannot determine the country a visitor is coming from. All hosts without a name will show up as Unresolved in the country list. Note: Sometimes, systems are intentionally not configured in the DNS, so a percentage of up to 30% for unresolved IP numbers is absolutely normal. The country report shows up in the Main window.

## Appendix C

### Definitions from ED's Implementation of WebTrends

**Hit** – An action on the Web site, such as when a user views a page or downloads a file

**Page View** – also called Page Impressions. Hit to HTML pages only; access to non-HTML documents are not counted.

**Visitor Session** – a session of activity (all hits) for one user of a web site. A unique user is determined by the IP address or cookie. (ED does not use cookies to identify unique users for logs.) By default, a user session is terminated when a user is inactive for more than 30 minutes. This duration can be changed from General panel in the Options, Web Log Analysis dialog. This is synonymous to a visit.

**Unique Visitor** – a unique IP address for the period of the report; may be authenticated using domain names or cookies (ED does not use the latter approach).

P:\CENDI\Working Groups\User Education\Metrics & Eval TG\Metrics & Eval Report - Final.DOC