the NATIONAL OCEAN SERVICE

Performance Measures and Web Analytics for Educators

NOS: HEALTHY COASTS, HEALTHY ECONOMIES

Delivering Educational Materials: The old way

Inside old Litchfield class 01

Delivering educational materials: The New Way (1970)

Web: the ultimate educational distraction

- About 5 percent of texts have been digitized (2006)
- Virtually all recorded music has been digitized
- About 12 percent of the half million films listed on IMDB have been digitized on DVD (2011)
- Number of Web sites on the Internet: 3 billion
- Number of blogs on the Internet: 700 million
- Videos watched on YouTube every day: 2 billion (2011)
- Every minute 45 hours of video uploaded to YouTube
- Number of images uploaded to Flickr every minute: 3,000 (2011)
- Number of images on Flickr, Sept. 2010: 5 billion
- Number of images uploaded to Facebook, every year: 36 billion (2011)

What makes the Web different

- It is random access (you don't have to start at the beginning)
- It is on demand (you don't have to wait for the 7:00 p.m. news, or the morning paper)
- It comes to you (you don't have to go to the theater, or a library)
- **It is indexed.**
- It is recorded, and you can get metrics

You can track activity on the Web. Sorta

- Web content is served by computers, and can be logged
- Because traffic is logged, there is the potential for tracking individuals
- Because there are next to no commercial privacy laws, corporations invest heavily in tracking Web visitors
- Because U.S. citizens are wary of their government, **they object to being tracked by the government**
- The Executive Office of the President, Office of Management and Budget, insists on privacy controls for Web tracking
- With few exceptions, government Web sites are prohibited from logging information which can be used for personal identification
- Additional restraints are in place for **sites targeting children**

What does a Web server log?

- Date and time
- What visitor requested
- Where visitor came from (referring page)
- Browser type
- Bytes transferred to visitor
- Numeric address (sorta) of your computer
- Status code (success, various types of failure)
- Individuals -- but only if the site requires a log-in

Typical Web log

- **125.sub-174-253-210.myvzw.com - - [01/Apr/2011:01:48:37] "GET http://www.papahanaumokuakea.gov/styles/styles_main.css HTTP/1.1" 304 240 "http://www.papahanaumokuakea.gov/resource/welcome.html" "Mozilla/5.0 (iPhone; U; CPU iPhone OS 4_2_6 like Mac OS X; en-us) AppleWebKit/533.17.9 (KHTML, like Gecko) Version/5.0.2 Mobile/8E200 Safari/6533.18.5" www.papahanaumokuakea.gov**

- On April 1, 2011, at 1:48 UT, someone on an iPhone used Verizon's network to look at the Papahanaumokuakea Web site, specifically the CSS style sheet for the Resource Protection page.

Web site lies

.125.sub-174-253-210.myvzw.com - - [01/Apr/2011:01:48:37] "GET http://www.papahanaumokuakea.gov/styles/styles_main.css HTTP/1.1" 304 240 "http://www.papahanaumokuakea.gov/resource/welcome.html" "Mozilla/5.0 (iPhone; U; CPU iPhone OS 4_2_6 like Mac OS X; en-us) AppleWebKit/533.17.9 (KHTML, like Gecko) Version/5.0.2 Mobile/8E200 Safari/6533.18.5" www.papahanaumokuakea.gov

- The address of a Web site visitor *rarely* points to an individual machine or device
- The Web browser identity is frequently obscured or modified

Example information from web log

The following examples are taken from a traditional Web log.

You can find a wealth of web metrics (all for National Ocean Service and its partners) at:

https://webstats.nos.noaa.gov/

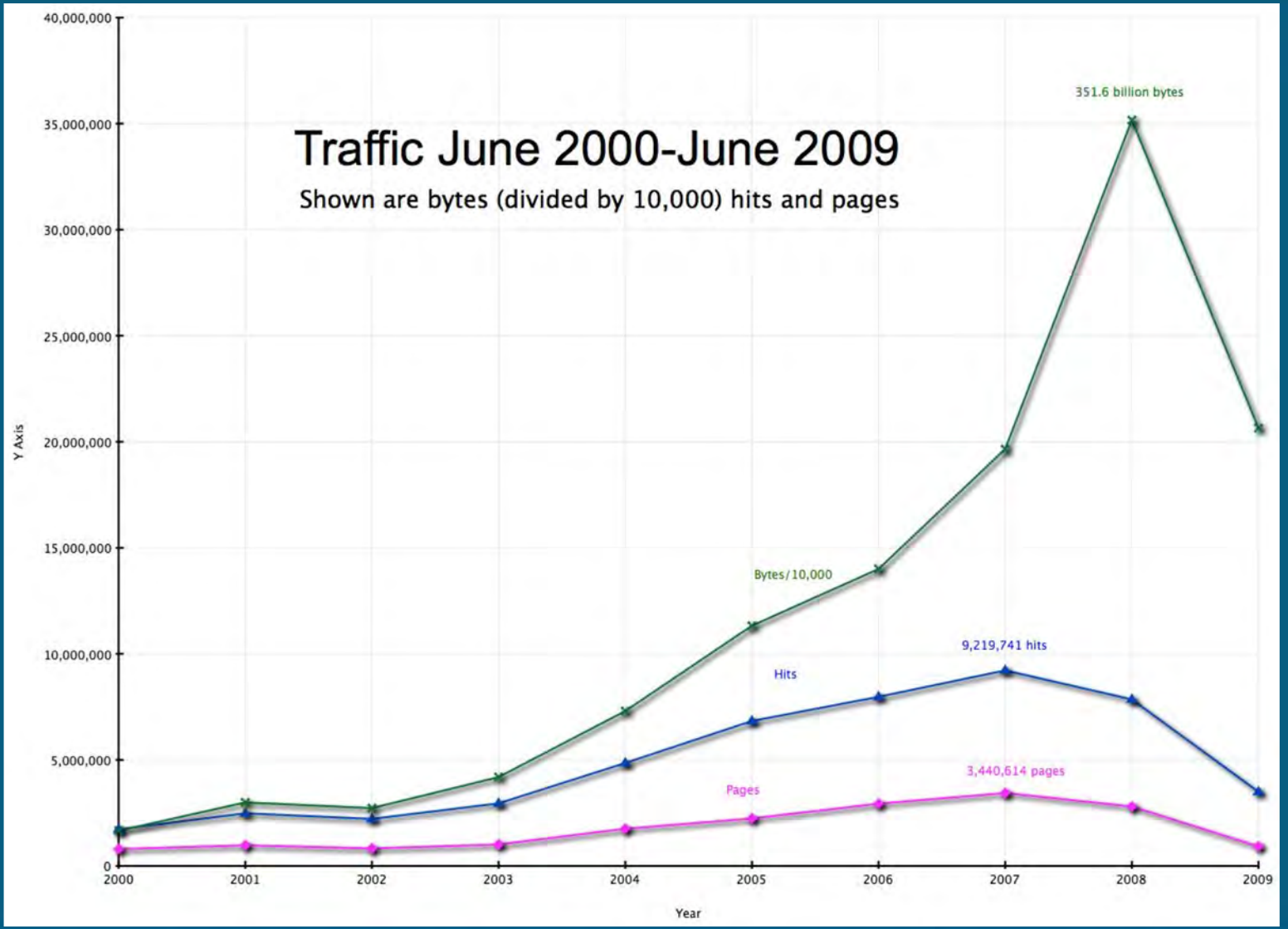
Use your NOAA credentials to log in.

Metrics for content providers

Time	Referrers	Content	Visitors	Visits	Problems
Hourly	Domains	Pages	Countries	Entry Point	Bad Links
Daily	Referrers	Downloads	Visitors	Exit Point	Failed Refers
Weekly	Search Engines	Graphics	Browsers	1 Page Visits	Gaps in Service
Monthly	Search Words	Other Requests	Computers	Steps	Hijacking
Quarterly	Search Phrases	All Requests	Robots	X per Y	Failed Requests
Yearly	To/From	by Group	Auth Users	Recent	Explanations
Cyclic Totals		by Directory		Duration	Least Requested
		by Type		View Time	Reloads

General	Bandwidth	Paths	Custom	Details	
Overview	Peak	by Source	Ref. Domains	Log Fields	
Program Status	by Request	by Destination		Local Search	
Servers	by Directory	Paths		Domains	
	by Type				
	Transfers				

Classic metrics: hits, pages, bytes



Overall View

How has traffic varied from month to month?

Date ?	Pages ?	Hits ?	Errors ?	GigaBytes ?	Visits ?	Visits ?
January 2010	692,510	11,000,291	89,139	187.5G	298,705	
February 2010	463,474	7,979,491	64,283	157.7G	257,390	
March 2010	507,788	10,336,523	98,802	174.0G	316,992	
April 2010	618,181	12,488,408	111,158	697.0G	376,612	
May 2010	721,829	14,462,099	424,621	421.5G	456,246	
June 2010	586,036	14,052,581	389,733	414.6G	564,942	
July 2010	407,033	10,330,041	114,245	252.9G	446,855	
August 2010	361,823	9,277,253	89,743	229.6G	304,473	
September 2010	505,169	12,773,929	112,367	281.2G	351,523	
October 2010	476,925	11,039,389	148,055	201.3G	346,641	
November 2010	408,327	9,496,585	94,908	212.2G	309,144	
December 2010	367,054	5,713,594	271,099	149.0G	445,848	

Pages over time

How often have pages been requested each of the last fourteen months?

Request ?	Pages ?	Dec 2010	Nov 2010	Oct 2010	Sep 2010	Aug 2010	Jul 2010	Jun 2010	May 2010	Apr 2010	Mar 2010	Feb 2010	Jan 2010
/	1,400,897	80,735	97,528	120,975	148,169	100,346	109,333	143,046	141,768	131,529	113,511	93,882	120,075
/robots.txt	208,445	14,174	11,561	15,050	14,140	12,494	14,816	18,236	19,370	21,387	20,184	21,453	25,580
/education/	74,850	4,324	4,857	5,979	6,118	3,961	5,641	7,588	9,155	7,278	6,519	5,450	7,980
/facts/oceandepth.html	58,991	3,880	3,421	4,640	4,771	3,789	3,653	3,938	8,355	15,077	1,827	1,607	4,033
/deepwaterhorizon /images.html	44,724	481	664	922	1,262	1,166	2,768	18,979	18,482	0	0	0	0
/dataexplorer/	37,712	2,539	2,889	3,429	3,592	2,772	3,413	3,884	4,345	3,040	2,670	2,313	2,826
/redirect.php	37,547	1,249	4,243	1,811	2,448	4,247	5,114	5,451	4,256	3,590	2,178	1,566	1,394
/about /supp_commerce_goal.html	32,467	85	93	97	152	168	2,111	12,861	15,657	1,020	88	70	65
/facts/	32,290	2,392	2,276	2,863	2,384	2,100	1,285	1,881	3,711	3,721	3,350	2,623	3,704
/education/kits/tides/	29,718	2,477	2,054	1,685	1,912	690	761	1,417	2,182	4,902	3,962	3,010	4,666
/education/kits/tides/media /supp_tide06a.html	26,727	1,699	2,235	1,813	876	506	933	1,791	4,429	4,234	2,817	2,791	2,603
/education/kits/tides /tides01_intro.html	24,784	911	951	1,036	1,255	794	961	1,149	1,890	4,791	3,515	3,086	4,445
/rss.html	23,876	549	568	639	563	562	620	2,089	4,015	4,066	3,289	3,009	3,907
/education/kits/currents/	23,048	887	1,342	1,171	1,387	654	615	1,106	2,596	2,865	3,424	2,629	4,372
/redtide /supp_noaareleases.html	22,968	2,742	2,138	2,196	1,953	1,639	1,733	1,927	2,481	2,190	1,707	863	1,399
/widgets/oceanfacts.html	22,939	1,238	1,427	2,352	1,797	2,479	4,744	1,756	1,916	1,579	1,123	871	1,657
/products/pubs_hypox.html	22,320	958	1,265	1,562	1,840	1,329	1,550	1,877	2,846	3,247	2,218	1,649	1,979
/education/kits/tides /tides02_cause.html	21,177	955	939	1,138	1,014	537	674	841	1,429	3,635	3,244	2,760	4,011
/facts/howmanyoceans.html	20,985	1,148	857	1,864	2,842	1,751	1,215	1,433	1,773	2,671	1,583	1,627	2,221
/kids/	20,951	934	1,228	1,591	2,437	929	1,053	1,670	2,870	2,364	1,859	1,843	2,173
/images.html	20,511	1,536	1,257	1,405	1,792	1,466	1,606	2,729	2,870	3,065	1,296	694	795
/topics/navops/marinenav/	20,473	656	551	726	729	677	1,383	2,343	2,906	3,235	2,214	1,980	3,073
/education/kits/corals/	20,284	982	931	1,213	1,006	564	623	1,128	1,990	3,936	2,879	1,762	3,270
/education/stories/lionfish/	19,627	7,416	925	1,147	963	867	833	838	1,707	1,325	1,557	1,043	1,006
/education/stories/oilymess/	19,538	741	939	999	1,340	999	1,798	2,499	5,371	1,045	2,089	606	1,112
/websites/retiredsites /supp_sotc_retired.html	19,508	1,229	957	1,264	1,514	980	1,260	1,784	2,043	2,614	2,100	1,710	2,053

Visitors

Which second level domains did the most requests come from?

Second Level Domain ?	% of Pages ?	Pages ?	Hits ?	GigaBytes ?
(Unknown)	19.22%	1,175,548	17,213,356	722.7G
noaa.gov	12.45%	761,283	57,571,873	344.7G
googlebot.com	8.27%	505,963	1,473,448	53.6G
search.msn.com	4.04%	247,256	400,242	67.7G
other.yahoo.net	3.99%	243,991	420,133	72.0G
comcast.net	3.90%	238,481	5,969,050	216.9G
verizon.net	3.46%	211,789	4,320,797	244.9G
scoutjet.com	2.19%	133,672	146,616	3.4G
rr.com	2.10%	128,620	2,563,282	103.9G
lexisnexis.com	2.03%	124,151	169,791	14.8G
cox.net	1.26%	76,917	2,749,982	77.3G
sbcglobal.net	1.21%	73,943	1,494,182	56.9G
musc.edu	1.10%	67,231	4,942,221	21.4G
bellsouth.net	0.92%	56,543	1,514,224	66.3G
charter.com	0.79%	48,072	983,053	47.1G
amazonaws.com	0.78%	47,733	111,665	10.4G
baidu.com	0.77%	47,297	50,089	1.7G

Downloads over time

How often have downloads been requested each of the last fourteen months?

Request ?	Hits ?	Dec 2010	Nov 2010	Oct 2010	Sep 2010	Aug 2010	Jul 2010	Jun 2010	May 2010	Apr 2010	Mar 2010	Feb 2010	Jan 2010
/education/tutorial_currents /lessons/currents_tutorial.pdf	193,824	18,088	48,911	31,663	26,250	8,918	9,288	17,959	32,747	0	0	0	0
/education/tutorial_tides/lessons /tides_tutorial.pdf	144,595	59,303	14,244	13,574	14,344	6,360	7,992	12,994	15,784	0	0	0	0
/websites/retiredsites/sotc_pdf /POP.PDF	113,987	7,746	9,191	11,900	13,995	6,347	5,572	7,519	11,985	11,855	10,009	8,175	9,693
/programs/mb/pdfs /coastal_pop_trends_complete.pdf	70,627	4,807	7,369	6,815	7,920	4,813	4,135	4,495	6,399	7,837	5,543	5,236	5,258
/education/tutorial_estuaries /lessons/estuaries_tutorial.pdf	53,095	5,727	12,353	9,363	7,127	4,412	5,222	6,261	2,630	0	0	0	0
/education/tutorial_corals/lessons /corals_tutorial.pdf	47,406	3,282	6,324	6,890	5,765	3,539	4,437	7,824	9,345	0	0	0	0
/education/for_fun /BuildUnderwaterRobot.pdf	38,300	3,429	2,720	3,835	4,542	2,853	2,854	2,834	3,323	2,910	2,286	2,600	4,114
/education/pd/climate/factsheets /howhuman.pdf	36,877	2,704	4,294	4,896	3,577	1,853	1,693	2,294	3,463	3,466	3,458	2,373	2,806
/education/for_fun /EndangeredSpeciesOrigami.pdf	32,669	4,306	2,090	2,480	2,285	2,227	2,106	3,534	4,537	3,276	1,839	2,093	1,896
/education/for_fun /BuildyourownWeatherStation.pdf	29,913	1,986	2,406	2,540	2,901	1,880	2,936	2,137	2,987	2,842	2,075	2,077	3,146
/education/for_fun /TornadoBottle.pdf	29,906	2,319	2,137	2,739	2,112	1,104	1,199	1,638	3,871	4,031	2,701	2,929	3,126
/education/for_fun /Makeanediblecoralreef.pdf	28,145	2,170	1,590	1,901	2,335	1,593	968	2,693	3,614	3,829	2,422	3,026	2,004
/websites/retiredsites/sotc_pdf /hab.pdf	26,999	1,783	2,032	3,130	2,240	1,169	1,705	2,350	3,043	3,338	2,827	1,481	1,901
/education/yos/curriculum /project_flow.pdf	25,756	1,990	2,836	2,605	4,384	1,215	631	1,272	4,598	2,352	1,517	1,198	1,158

What were people searching for?

Which search phrases were used to locate the site?

% of Visits ?	Visits ?	Goal Ones % of Visits ?	Goal Ones ?	Avg. Steps ?	$ Value ?	Phrase ?
5.46%	29,339	0.00%	0	1.2	0.00	how deep is the ocean
1.92%	10,301	0.00%	0	1.1	0.00	how many oceans are there
1.53%	8,196	0.00%	0	1.8	0.00	lionfish invasion
1.50%	8,055	0.00%	0	0.8	0.00	gulf of mexico oil spill
0.84%	4,503	0.00%	0	1.1	0.00	why is the ocean blue
0.76%	4,083	0.00%	0	3.1	0.00	ocean facts
0.74%	3,979	0.00%	0	0.7	0.00	gulf oil spill
0.73%	3,937	0.00%	0	0.9	0.00	hail
0.72%	3,880	0.00%	0	0.8	0.00	evaporation
0.63%	3,365	0.00%	0	0.8	0.00	supercell
0.58%	3,124	0.00%	0	3.3	0.00	national ocean service
0.48%	2,569	0.00%	0	1.4	0.00	noaa
0.47%	2,527	0.00%	0	1.2	0.00	oceanographer
0.45%	2,406	0.00%	0	0.8	0.00	oil spill in gulf of mexico
0.44%	2,372	0.00%	0	0.9	0.00	thermocline
0.42%	2,276	0.00%	0	2.3	0.00	tides
0.42%	2,245	0.00%	0	1.1	0.00	how many oceans
0.41%	2,220	0.00%	0	2.5	0.00	nos
0.39%	2,121	0.00%	0	0.7	0.00	oil spill gulf of mexico
0.39%	2,104	0.00%	0	1.0	0.00	pillow lava
0.38%	2,063	0.00%	0	0.9	0.00	cyclones
0.37%	1,988	0.00%	0	3.1	0.00	estuaries
0.35%	1,900	0.00%	0	1.6	0.00	zooxanthellae

What is not logged

- Visitor name, age, gender, etc.
- What was the visitor looking for?
- Did the visitor find it?
- Was the visitor satisfied with what was found?
- Did the visitor *trust* what was found?
- How will the visitor use what was found?
- Did the visitor have more confidence in some other site?
- Did the visitor read what you wrote?

What was deliberately not reported

Disable Reports

Time	Content	Referrers	Visitors	Browser
☐ Hourly Report	☐ Pages	☐ Referring Domains	☐ Countries	☐ Browsers
☐ Hourly Metrics	☐ Over Time	☐ by Steps	☐ by GeoIP	☐ Versions
☐ Daily Report	☐ Downloads	☑ by Value	☐ Regions	☐ by Platform
☐ Daily Metrics	☐ Over Time	☑ by Interest	☐ by GeoIP	☐ Agents
☐ Weekly Report	☐ Graphics	☐ Referrers	☐ States	☐ Screen Size
☐ Weekly Metrics	☐ Over Time	☐ Over Time	☐ All Domains	☑ Color Depth
☐ Monthly Report	☐ Other Requests	☐ Full	☑ Hosts	☑ Window Width
☐ Monthly Metrics	☐ Over Time	☑ New	☑ with IP	☑ Window Height
☑ Quarterly Report	☐ All Requests	☐ Search Words	☐ Known Robots	☐ Javascript
☑ Quarterly Metrics	☐ Over Time	☑ by Value	☐ Crawl Dates	☑ Java
☐ Yearly Report	☑ New	☐ Search Phrases	☐ by Indexer	☑ Cookies
☐ Yearly Metrics	☐ Directories	☑ by Goal Ones	☑ Hosts	☑ Plugins
☐ Time of Day	☑ Groups	☑ by Value	☐ Modem Speed	☑ Language
☐ Quarter of Day		☑ New	☐ Computers	☑ Language Codes
☐ Day of the Week		☐ Entry by Phrase	☐ Platforms	
☐ Month of the Year		☐ Phrase by Entry	☐ Peak Days	

Problems	Visit Details	Subsets
☐ Bad Links	☐ Visit Duration	☑ Goal Ones
☐ with Requests	☐ Pages per Visit	☑ Goal Twos
☐ by Request	☐ Hits per Visit	☑ Featured
☐ Failed Requests	☐ Errors per Visit	☑ User One
☐ Over Time	☐ Bytes per Visit	☑ User Two
☐ Explanations	☐ Uniques per Visit	☑ User Three
☐ Gaps in Service	☑ Visits per Host	☑ User Four
☐ Least Requested		☑ User Five
☐ Reloads		☑ User Six
☐ Failed Refers		☑ User Seven
☐ by Domain		☑ User Eight
☐ with Requests		☑ User Nine
☐ by Request		☑ User Ten
		☑ User Eleven
		☑ User Twelve

☐ Details	☐ Referred From	☑ Reqs by User	☑ Words by Goal2
	☐ Complete	☑ Users by Req	☐ Protocol
	☑ Images loaded	☑ Cookies	☐ Servers
	☑ Where Found	☐ File Types	☐ RefDom by Bytes
	☑ First Image	☑ RefDom by Goal	☑ Hosts by Bytes
	☐ Fan In/Out	☐ Local Words	☐ Robots by Bytes
		☐ Local Phrases	☐ Domains by Bytes
		☑ Campaigns	☑ TrackTag by Goal2
		☑ by Goal Ones	☑ RefDom by Goal2
		☑ by Value	☑ Phrases by Goal2

☐ Pages	☐ Pages	☑ Quarter Day
☐ Downloads	☐ Downloads	☑ 1/4 Day Metrics
☐ Graphics	☐ Graphics	☐ Compare
☐ Others	☐ Others	☐ Hourly Visits
☐ Requests	☐ Requests	☑ 1/4 Day Visits
☐ Refers	☐ Refers	☐ Daily Visits
☐ Phrases	☐ Phrases	☐ Weekly Visits
☐ Directories	☐ Directories	☐ Monthly Visits
☑ Campaigns	☑ Campaigns	☐ Quarterly Visits
☐ Search Engines	☐ Search Engines	☐ Yearly Visits
	☑ Featured	

Google Analytics and DOC

- Google Analytics is a free (or paid) Web service offered by Google
- DOC classifies Google Analytic as a social media service: http://ocio.os.doc.gov/ITPolicyandPrograms/Policy___Standards/PROD01_009476
- Google analytics is also considered a "web measurement" service, so is covered by another policy as well: http://ocio.os.doc.gov/ITPolicyandPrograms/Policy___Standards/PROD01_009664
- Because Google Analytics is a third party service that can be used to track individuals, it is also covered by the DOC Privacy Policy Statements and Information Collection policy: http://www.osec.doc.gov/webresources/policies/policy7a_privacy_notices.htm

Google Analytics and DOC (cont.)

Google Analytics basics:

- A small piece of code is **embedded on a Web page**
- When a user visits, that code **contacts Google**
- A larger piece of code (18K) is **downloaded to the visitor's browser**
- That larger piece of code collects information and **sends it back to Google**
- That larger piece of code also **sets a Web cookie on the visitor's machine**
- The Web bug is used to **track that user's data on *all* sites** that use Google Analytics
- Result: using Google Analytics on a NOAA site essentially means you are helping Google track user behavior not only on commercial sites but *also* on government sites

Google Analytics and NOAA

- Google Analytics requires that accounts be tied to a valid E-mail address
- DOC and NOAA, for security purposes, require use of official E-mail addresses for official functions
- DOC approved Google Analytics accounts, therefore, must be tied to official E-mail addresses
- On Sept. 1, 2011, NOAA moved to Google Mail
- On Sept. 1, 2011, NOAA blocked use of Google Analytics accounts (as well as YouTube admin accounts, Google Webmaster Tools accounts, Google Maps admin accounts, Google Feedburner accounts, etc.) tied to noaa.gov addresses
- Therefore: **you cannot use Google Analytics in noaa.gov until this is resolved**

Google Analytics and NOAA (cont.)

- Because Google Analytics is classified as social media by DOC, you must apply for its use: https://socmed.my.commerce.gov/
- Each NOAA line office has social media IT communications contacts for social media accounts. Accounts must be approved by both before they are sent on for NOAA for approval, then an additional clearance is required at the Commerce level.
- As Google Analytics tracks activity of visitors by (a) placing a Web bug on a government page, (b) downloading code from a non-government site, and (c) placing a cookie on a visitor's machine, every site that contemplates using Google Analytics needs a PIA (Privacy Impact Assessment). This requires work by an IT security specialist, the server's system administrator, and the site content manager
- Sites that can be reasonably assumed to be aimed at children will require additional review

Google Analytics limitations

- Google Analytics works by placing a Web bug on a page
- Result: Google Analytics can only track *pages*
  - Downloads cannot be tracked
  - Video and sound files cannot be tracked
  - Failed requests cannot be tracked
  - .zip and other archives cannot be tracked
  - PDFs cannot be tracked
  - Bandwidth cannot be tracked
  - In fact, most things you put on a Web site can't be tracked
- Google Analytics can also be disabled if:
  - JavaScript is disabled on the client
  - The user has a plug-in explicitly for blocking Google Analytics or ad blocking plug-ins or script blocking plug-ins
  - The user blocks cookies
  - The user clears cookies

Google Analytics and NOAA (cont.)

- Use of Google Analytics will require extensive modification of the site **Privacy Policy**, describing in full the use of Google Analytics as a tracking mechanism
- The Executive Office of the President, Office of Management and Budget, also requires that sites using such tracking tools offer visitors an easy way of **turning off tracking**.
  - The most commonly proposed method is suggesting that visitors turn off JavaScript
  - Turning off JavaScript "breaks" most modern sites, disables all AJAX functions, and may render sites unusable
  - Because Google Analytics is enabled through a Web bug embedded in the page, sites must be very carefully designed to work without JavaScript

Google Analytics limitations (cont.)

- Many Google reports are based on sampling
- Sampling errors are particularly large for low volume sites
- Sometimes, Google results are simply bizarrely incorrect

Demonstration

- National Ocean Service:
- Site Metrics:
  - https://webstats.nos.noaa.gov/sum2012/oceanservice/menu.html
- Education Site Metrics:
  - https://webstats.nos.noaa.gov/sum2012/education/menu.html
- Google comparative metrics (non-NOAA site):
  - http://www.wap.org/metrics/
  - http://www.google.com/analytics/

Summary

- If you want accuracy and completeness, traditional log-based Web metrics are best.
- Google Analytics are a good choice when you do not have access, for whatever reason, to log-based metrics.
- Google Analytics are *best* used *in addition* to traditional log-based Web metrics.
- Significant challenges are posed in using Google Analytics for education sites (particularly government sites) because the techniques used by Google stand a much higher chance of invasive tracking of children.
- If you do use Google Analytics, make sure you follow all the rules and properly apply for the account through DOC, do a Privacy Impact Assessment of your site (first), and clearly spell out the use of Google Analytics in your Privacy Policy. You may *not* use the Privacy Policy on www.noaa.gov as it is inadequate.

Contact

Lawrence I. Charters
National Ocean Service
National Oceanic and Atmospheric Administration
U.S. Department of Commerce
lawrence.charters@noaa.gov