

Generating Site Statistics

The Health and Retirement Study

Introduction

Web site usage statistics are maintained for [the main HRS Web site](#) and the [Participant Web site](#). Summary statistics for the main site are available to the [general public](#); detailed statistics are available only to HRS staff and are displayed on the [Staff Web site](#). This document describes how to generate and maintain usage statistics for public and staff use.

Statistics for the Main HRS Web Site

Public View

HRS began to collect usage statistics when the public site was established in 1995. Since then, usage information has been collected on key elements of the site on a monthly basis. A 12-month detailed record is supplemented by a yearly summary table.

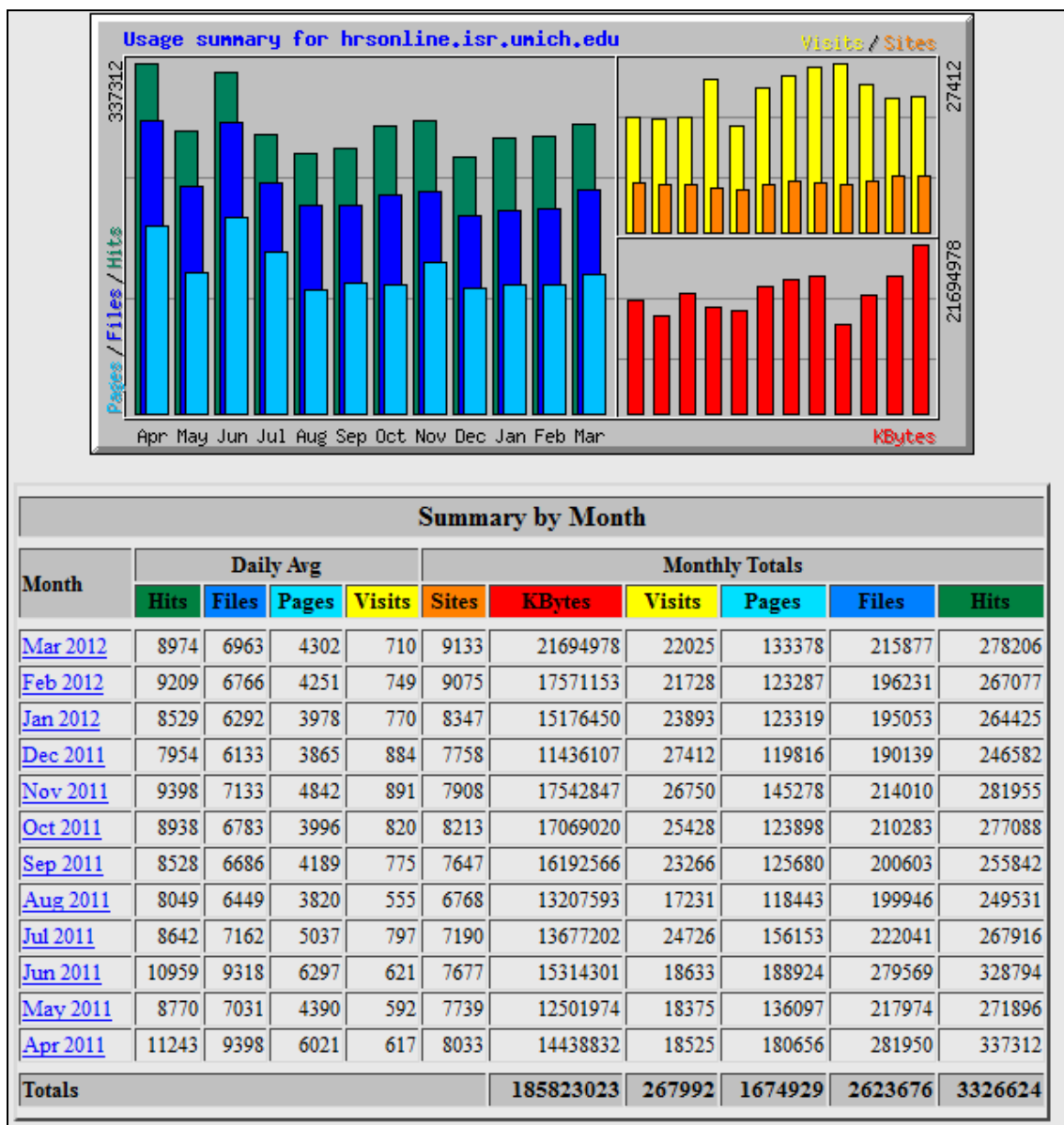
12 Month Data Summary									
Year/Month	Visits		User Registrations		Concordance Searches		Bibliography Searches		Notes
	Month	Cumulative	Month	Cumulative	Month	Cumulative	Month	Cumulative	
Mar 2012	13,883	429,387	229	15,574	2,274	84,785	682	20,343	HRS genetic data posted to dbGAP
Feb 2012	14,320	415,504	272	15,345	3,074	82,511	806	19,661	
Jan 2012	16,157	401,184	175	15,073	1,979	79,437	845	18,855	
Dec 2011	15,504	385,027	133	14,898	1,488	77,458	866	18,010	
Nov 2011	15,136	369,523	187	14,765	3,173	75,970	731	17,144	
Oct 2011	15,484	354,387	259	14,578	3,401	72,797	827	16,413	2010 Exit (Final) Released
Sep 2011	13,665	338,903	227	14,319	3,083	69,396	864	15,586	
Aug 2011	11,428	325,238	165	14,029	2,270	66,313	521	14,722	
Jul 2011	11,515	313,801	158	13,927	2,200	64,043	581	14,201	2010 Core (Early) Released
Jun 2011	12,325	302,295	172	13,769	2,254	61,843	707	13,620	
May 2011	12,362	289,970	169	13,597	2,463	59,589	682	12,913	
Apr 2011	12,575	277,608	187	13,428	3,109	57,126	615	12,231	

- The *Visits* column shows the number of **visits** to the site. A visit is initiated when a remote site requests a page on the HRS server for the first time. All requests from that site within a given time period are considered to be part of the same visit. If the site makes a request to your server, and the timeout value has been reached (default is 30 minutes), a new visit is started and counted.
- The *User Registrations* column reflects the number of new usernames submitted as part of the registration process for access to HRS public data. Note: The current file download system was introduced on 1 January 2001. Registration is mandatory for all public data users.
- The *Concordance* column shows the total number of **searches**; i.e., calls to the Concordance search engine (`/cgi-bin/getconca.pl`).

- The *Bibliography* column shows the total number of **search** calls to the Bibliography search engine (/cgi-bin/getbib.pl). Additional summary information specific to the Bibliography (e.g., [counts by workform and content area](#)) is available on the [Papers and Publications](#) section of the main site.

Staff View

Additional information concerning the main HRS site is available through the [Usage Statistics](#) page of the Staff site. This information is used to create summary information displayed to the public using the procedures described below. The detailed information on the Staff site includes graphs and tables that display monthly, daily and hourly summaries of Web site use. Information is also available on which site resources are used most frequently, top referrers, and most frequent visitors.



Summary by Month

Month	Daily Avg				Monthly Totals					
	Hits	Files	Pages	Visits	Sites	KBytes	Visits	Pages	Files	Hits
Mar 2012	8974	6963	4302	710	9133	21694978	22025	133378	215877	278206
Feb 2012	9209	6766	4251	749	9075	17571153	21728	123287	196231	267077
Jan 2012	8529	6292	3978	770	8347	15176450	23893	123319	195053	264425
Dec 2011	7954	6133	3865	884	7758	11436107	27412	119816	190139	246582
Nov 2011	9398	7133	4842	891	7908	17542847	26750	145278	214010	281955
Oct 2011	8938	6783	3996	820	8213	17069020	25428	123898	210283	277088
Sep 2011	8528	6686	4189	775	7647	16192566	23266	125680	200603	255842
Aug 2011	8049	6449	3820	555	6768	13207593	17231	118443	199946	249531
Jul 2011	8642	7162	5037	797	7190	13677202	24726	156153	222041	267916
Jun 2011	10959	9318	6297	621	7677	15314301	18633	188924	279569	328794
May 2011	8770	7031	4390	592	7739	12501974	18375	136097	217974	271896
Apr 2011	11243	9398	6021	617	8033	14438832	18525	180656	281950	337312
Totals						185823023	267992	1674929	2623676	3326624

Selecting Statistics for Public Display

Information on Home Page visits

Go to the [Usage Statistics](#) page of the Staff web site, select the current month and then page down to the following table. Save this information for use in updating the public view table.

Top 10 of 4086 Total Entry Pages					
#	Hits		Visits		URL
1	34129	1.57%	6081	38.58%	/
2	1312	0.06%	179	1.13%	/rda/
3	1707	0.08%	152	0.96%	/docs/
4	913	0.04%	143	0.91%	/data/
5	1889	0.09%	131	0.83%	/intro/
6	649	0.03%	130	0.82%	/rda/userdocs/
7	1440	0.07%	117	0.74%	/news/
8	1437	0.07%	114	0.72%	/meta/
9	1370	0.06%	114	0.72%	/meta/alerts/
10	755	0.03%	110	0.70%	/papers/

Information on Bibliography and Concordance Searches

Page down until you find the Top 100 of Total URLs table. Then look for references to /cgi-bin/getconc2.pl and /cgi-bin/getbib.pl. Select the number of hits for each one.

Top 100 of 10433 Total URLs					
#	Hits		KBytes		URL
1	70926	25.49%	1520402	7.01%	/
2	12950	4.65%	146898	0.68%	/favicon.ico
3	12184	4.38%	38847	0.18%	/modules/news/rss/hrs.xml
4	11085	3.98%	1101416	5.08%	/modules/meta/rand/randhrsl/randhrsl.pdf
5	7213	2.59%	67444	0.31%	/includes/stylev2.css
6	4336	1.56%	5774	0.03%	/images/favicon.ico
7	2274	0.82%	130730	0.60%	/cgi-bin/getconc2.pl
8	2266	0.81%	247	0.00%	/robots.txt
9	1960	0.70%	22223	0.10%	/news/hrs.xml
10	1750	0.63%	166608	0.77%	/modules/meta/adams/desc/adams1dd.pdf
11	1561	0.56%	53860	0.25%	/modules/meta/2010/core/qnaire/online/HRS2010_SAO_Final.pdf
12	1387	0.50%	62533	0.29%	/sitedocs/dmgt/OverviewofHRSPublicData.pdf
13	1209	0.43%	112658	0.52%	/sitedocs/surveydesign.pdf
14	1136	0.41%	13216	0.06%	/includes/sortable.js
15	1110	0.40%	104173	0.48%	/modules/meta/rand/randhrsk/randhrsk.pdf
16	1010	0.36%	35403	0.16%	/concord
17	980	0.35%	88531	0.41%	/sitedocs/genetics/datasheet_omni_whole-genome_arrays.pdf
18	911	0.33%	44018	0.20%	/sitedocs/userg/dr-006.pdf
19	788	0.28%	58316	0.27%	/sitedocs/userg/HRSSAMP.pdf
20	682	0.25%	126599	0.58%	/cgi-bin/getbib.pl

Information on User Registrations

To get information on user registrations in the past month, go to the [Usage Statistics Page](#) on the Staff Web Site.

1. Click on the User Registrations Link

Home	Usage Statistics for HRS Web Sites.
Documentation	HRS Public Web Site (hrsonline)
Programming Resources	<ul style="list-style-type: none"> User Registrations by Year and Month Staff Version: Complete details on the last 12 months of site activity. These statistics are only available to internal ISR IP addresses. The database from which they are derived was initialized on January 31, 2006. Public Version: A summary of site activity based on Home Page hits and File Distribution System registrations.
Privacy	
Production	
Security	HRS Participant Web Site (hrsparticipants) Usage Statistics
Video/Voice/Technical	<ul style="list-style-type: none"> Staff Version: Initialized on September 1, 2007
Web Design	Bibliography Statistics
Web Site Statistics	<ul style="list-style-type: none"> Display Table: Cumulative Workform Counts Display Table: Interest Site Area Counts
SRC Computing Support	isr-hrsweb Usage Statistics
External Links	<ul style="list-style-type: none"> 2006: August ; July ; June ; May ; April ; March ; February ; January Previous Years: 2005; 2004; 2003; 2002

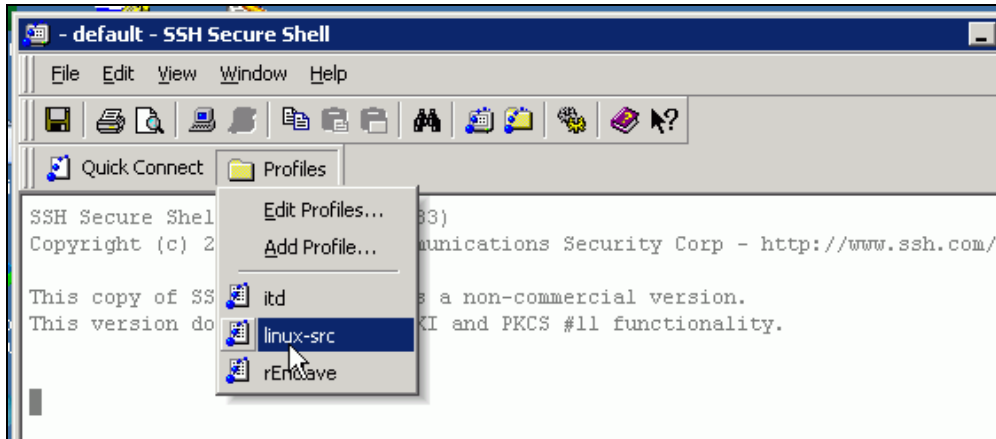
2. Get information for the year and month in question; for example, the new user count for June 2012 was 184.

Year	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec	Total
2001		1	5		147	24	24	31	19	25	35	30	341
2002	46	42	32	29	47	68	102	78	95	131	120	83	873
2003	100	127	126	99	90	90	75	75	105	121	82	86	1176
2004	102	134	124	120	85	102	71	71	111	123	105	69	1217
2005	123	123	80	163	103	115	75	103	103	116	98	49	1251
2006	90	112	110	121	92	116	70	96	105	141	119	76	1248
2007	97	119	119	135	99	160	110	92	105	128	138	85	1387
2008	122	100	98	121	104	95	102	83	131	137	106	74	1273
2009	94	143	118	145	104	101	79	78	141	157	117	76	1353
2010	129	137	165	153	137	141	131	130	133	169	190	136	1751
2011	163	154	202	165	149	156	150	161	209	256	181	143	2089
2012	167	265	228	204	170	184	50						1268
Total to date													15227

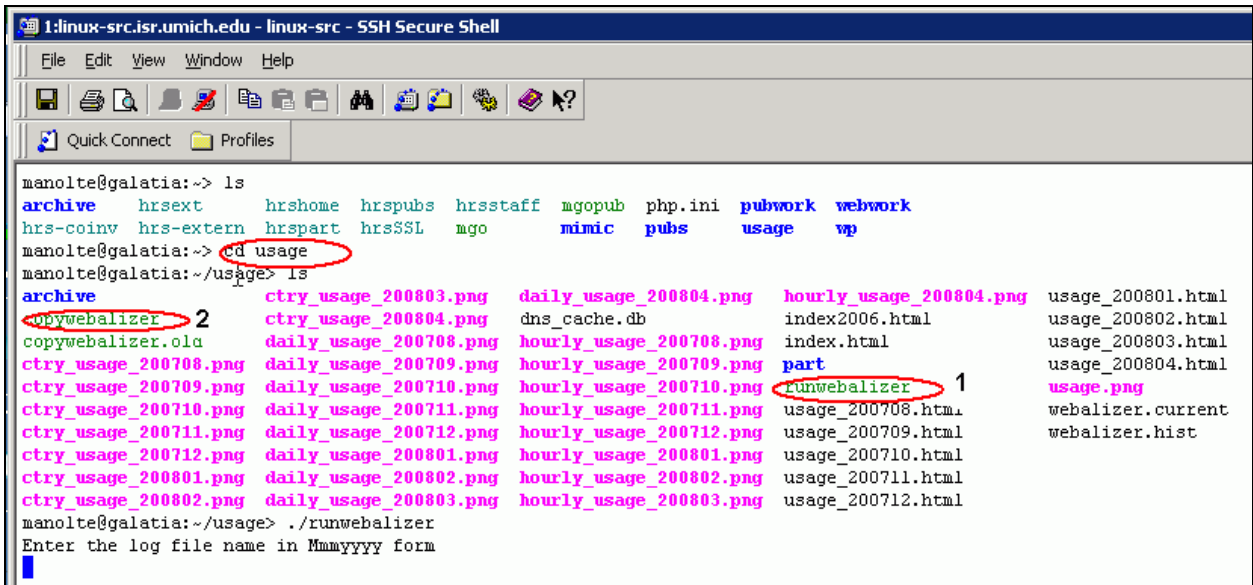
How to Generate the Information Described Above

With the exception of User Registrations, the usage tables described above are generated by a utility called [Webalizer](#) that uses as input the access and error logs produced during the normal operation of the Apache web server software. To run the monthly Webalizer updates, do the following:

1. Using SSH, open a command line link to the HRS web server



2. Navigate to the usage folder.



3. *Start the Webalizer process* (./runwebalizer -- **1** above). The logfile name should be specified as MM/YYYY, e.g. Mar2012. You may receive many warning messages of the type shown below, ignore them but make sure that the program terminates properly.

```
Warning: Truncating oversized referrer field [230281]
[new_snode] Warning: String exceeds storage size (70)
[new_snode] Warning: String exceeds storage size (64)
[new_snode] Warning: String exceeds storage size (112)
[new_snode] Warning: String exceeds storage size (73)
[new_snode] Warning: String exceeds storage size (73)
[new_snode] Warning: String exceeds storage size (76)
[new_snode] Warning: String exceeds storage size (76)
[new_snode] Warning: String exceeds storage size (76)
Warning: Truncating oversized referrer field [274934]
[new_snode] Warning: String exceeds storage size (71)
[new_snode] Warning: String exceeds storage size (213)
Saving current run data... [03/31/2012 23:59:43]
Generating report for March 2012
Generating summary report
Saving history information...
279241 records (1035 bad) in 197.00 seconds, 1417/sec
```

4. *When the Webalizer process is complete, copy the results to the Staff Web site* (./copywebalizer -- **2** above).

```
manolte@galatia:~/usage> ./copywebalizer
Enter the file suffix in YYYYMM form
201203

***Copying usage files to staff web site (2009) 201203
usage_201203.html copied

ctry_usage_201203.png copied

daily_usage_201203.png copied

hourly_usage_201203.png copied

old index.html backed up
index.html copied

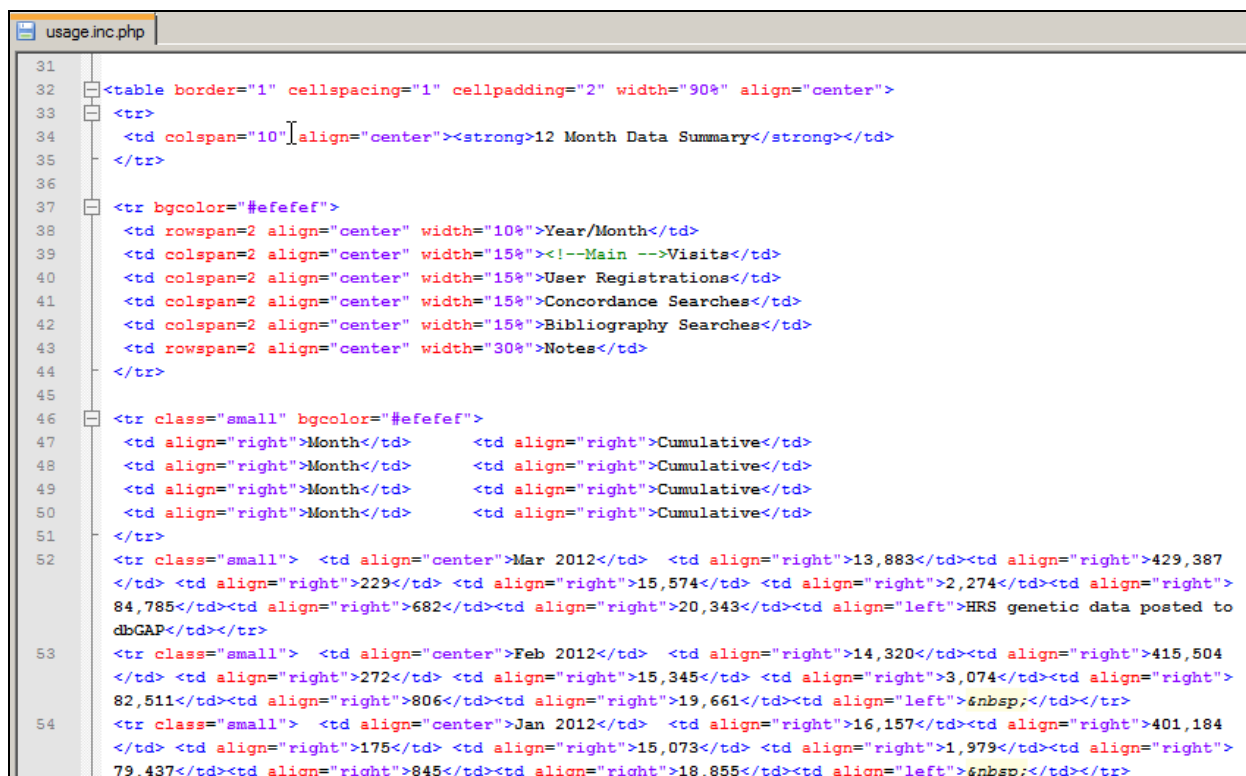
old usage.png backed up
usage.png copied

Done!
```

Updating the Usage Page

As explained elsewhere, much of the content on the HRS Web site is based on interactions between PHP code and the Apache Server. In the case of the Site Statistics page, all navigation information is added by a PHP program. This means that the only action required to maintain the page is to edit the HTML code.

Begin by opening `usage.inc.php`¹ in any convenient text editor. Scroll down until you reach the monthly portion of the usage table. Now enter the information that you collected using the techniques outlined above.



```
31
32 <table border="1" cellspacing="1" cellpadding="2" width="90%" align="center">
33 <tr>
34 <td colspan="10" align="center"><strong>12 Month Data Summary</strong></td>
35 </tr>
36
37 <tr bgcolor="#efefef">
38 <td rowspan="2" align="center" width="10%">Year/Month</td>
39 <td colspan="2" align="center" width="15%"><!--Main -->Visits</td>
40 <td colspan="2" align="center" width="15%">User Registrations</td>
41 <td colspan="2" align="center" width="15%">Concordance Searches</td>
42 <td colspan="2" align="center" width="15%">Bibliography Searches</td>
43 <td rowspan="2" align="center" width="30%">Notes</td>
44 </tr>
45
46 <tr class="small" bgcolor="#efefef">
47 <td align="right">Month</td> <td align="right">Cumulative</td>
48 <td align="right">Month</td> <td align="right">Cumulative</td>
49 <td align="right">Month</td> <td align="right">Cumulative</td>
50 <td align="right">Month</td> <td align="right">Cumulative</td>
51 </tr>
52 <tr class="small"> <td align="center">Mar 2012</td> <td align="right">13,883</td><td align="right">429,387
53 </td> <td align="right">229</td> <td align="right">15,574</td> <td align="right">2,274</td><td align="right">
54 84,785</td><td align="right">682</td><td align="right">20,343</td><td align="left">HRS genetic data posted to
dbGAP</td></tr>
53 <tr class="small"> <td align="center">Feb 2012</td> <td align="right">14,320</td><td align="right">415,504
54 </td> <td align="right">272</td> <td align="right">15,345</td> <td align="right">3,074</td><td align="right">
82,511</td><td align="right">806</td><td align="right">19,661</td><td align="left">&nbsp;</td></tr>
54 <tr class="small"> <td align="center">Jan 2012</td> <td align="right">16,157</td><td align="right">401,184
55 </td> <td align="right">175</td> <td align="right">15,073</td> <td align="right">1,979</td><td align="right">
79,437</td><td align="right">845</td><td align="right">18,855</td><td align="left">&nbsp;</td></tr>
```

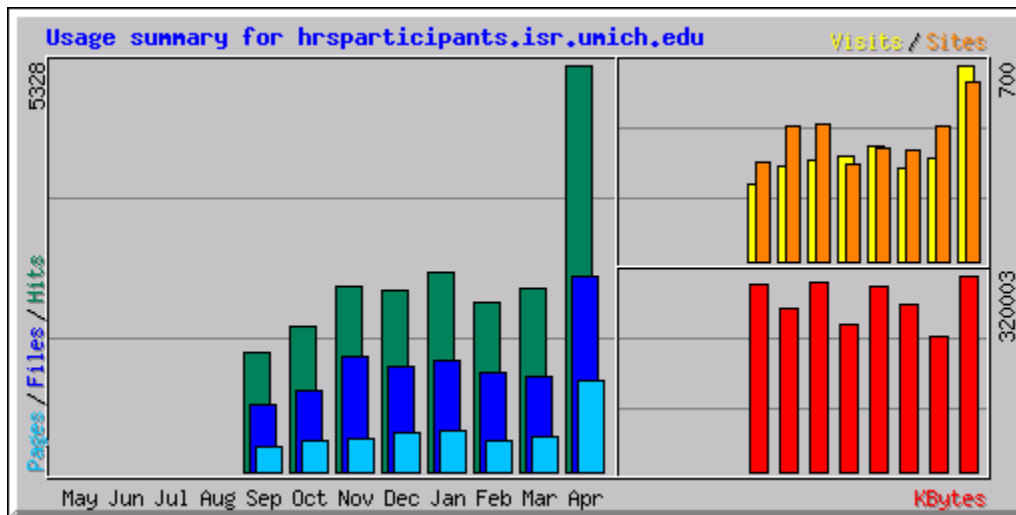
Add YTD information from the past month to the data items for the current month. Since we only want to display one year's worth of monthly data, the final step is to comment out the corresponding month from the previous year. At the end of each year, a new yearly summary row should be added to the table containing YTD information for the year.

¹ \\linux-src\www\vhosts\hrsonline.isr.umich.edu\html\modules\news\usage.inc.php

Participant Site

Staff View

At the present time, participant site statistics are not published for general use. However, information interaction between study participants and their Web site is available to HRS staff members through the [Web Site Usage Statistics Page](#).



Creating the Staff Web site usage tables

Participant site information is generated in much the same fashion as that described above.

```
ctry_usage_200801.png  daily_usage_200802.png  hourly_usage_200802.png  usage_200711.html
ctry_usage_200802.png  daily_usage_200803.png  hourly_usage_200803.png  usage_200712.html
manolte@galatia:~/usage> cd part
manolte@galatia:~/usage/part> ls
copywebalizer_part      ctry_usage_200804.png  dns_cache.db            index.html              usage_200804.html
copywebalizer_part_old.txt  daily_usage_200709.png  hourly_usage_200709.png  funwebalizer_part      usage.png
ctry_usage_200709.png    daily_usage_200710.png  hourly_usage_200710.png  usage_200709.html     webalizer.current
ctry_usage_200710.png    daily_usage_200711.png  hourly_usage_200711.png  usage_200710.html     webalizer.current.a
ctry_usage_200711.png    daily_usage_200712.png  hourly_usage_200712.png  usage_200711.html     webalizer.current.save
ctry_usage_200712.png    daily_usage_200801.png  hourly_usage_200801.png  usage_200712.html     webalizer.hist
ctry_usage_200801.png    daily_usage_200802.png  hourly_usage_200802.png  usage_200801.html     webalizer.hist.a
ctry_usage_200802.png    daily_usage_200803.png  hourly_usage_200803.png  usage_200802.html     webalizer.hist.save
ctry_usage_200803.png    daily_usage_200804.png  hourly_usage_200804.png  usage_200803.html
manolte@galatia:~/usage/part>
```

1. Using SSH, open a command line link to the HRS web server
2. Navigate to the usage folder
3. Start the Webalizer process (./runwebalizer_part2).
4. When the Webalizer process is complete, copy the results to the Staff Web site (./copywebalizer_part).

The command line dialog looks like this:

```
manolte@galatia:~/usage> cd participant/
manolte@galatia:~/usage/participant> ./runwebalizer_part2
Enter the log file name in Mmmyyyy form (e.g. Sep2007)
Mar2012
Webalizer V2.01-10 (Linux 2.6.16.60-0.69.1-smp) English
Using logfile /srv/www/htdocs/webalizer/archive/hrspart-access_log/Mar2012_log.gz (gzip-clf)
DNS Lookup (10):
870 addresses in 1944.50 seconds
Using DNS cache file dns_cache.db
Creating output in .
Hostname for reports is 'hrsparticipants.isr.umich.edu'
Reading history file... webalizer.hist
Reading previous run data.. webalizer.current
[new_snode] Warning: String exceeds storage size (66)
[new_snode] Warning: String exceeds storage size (234)
Saving current run data... [03/31/2012 23:20:49]
Generating report for March 2012
Generating summary report
Saving history information...
7227 records in 7.00 seconds, 1032/sec
manolte@galatia:~/usage/participant>
manolte@galatia:~/usage/participant>
manolte@galatia:~/usage/participant> ./copywebalizer_part
Enter the file suffix in YYYYMM form
201203

***Copying usage files to staff web site 201203
  usage_201203.html copied

  ctry_usage_201203.png copied

  daily_usage_201203.png copied

  hourly_usage_201203.png copied

  old index.html backed up
  index.html copied

  old usage.png backed up
  usage.png copied

Done!
manolte@galatia:~/usage/participant> █
```

APPENDIX: Command File Listings

1. ./runwebalizer

```
usage.inc.php  runwebalizer  copywebalizer

1  echo "Enter the log file name in Mmmmyyyy form"
2  read blah
3  if [ "$blah" = "" ]
4  then
5      echo "...the log file name should be something like 'May2006'"
6  else
7      webalizer -D dns_cache.db -U 100 -n hrsonline.isr.umich.edu -p -o .
      /srv/www/htdocs/webalizer/archive/hrs-access_log/$blah\_log.gz
8  fi
9
```

2. ./copywebalizer

```
usage.inc.php  runwebalizer  copywebalizer

1  echo "Enter the file suffix in YYYYMM form"
2  read blah
3  echo $blah
4  if [ "$blah" = "" ]
5  then
6      echo "...the log file name should be something like '200712'; you typed $blah"
7  else
8      echo ""
9      echo "***Copying usage files to staff web site (2009)" $blah
10
11     cp usage_$blah.html /srv/www/vhosts/hrs-staff.isr.umich.edu/html/webx/
12
13     echo "  usage_$blah.html copied"
14     echo " "
15
16     cp ctry_usage_$blah.png /srv/www/vhosts/hrs-staff.isr.umich.edu/html/webx/ctry_usage_$blah.png
17     echo "  ctry_usage_$blah.png copied"
18     echo " "
19
20     cp daily_usage_$blah.png /srv/www/vhosts/hrs-staff.isr.umich.edu/html/webx/daily_usage_$blah.png
21     echo "  daily_usage_$blah.png copied"
22     echo " "
23
24     cp hourly_usage_$blah.png /srv/www/vhosts/hrs-staff.isr.umich.edu/html/webx/hourly_usage_$blah.png
25     echo "  hourly_usage_$blah.png copied"
26     echo " "
27
28     tmp1="/srv/www/vhosts/hrs-staff.isr.umich.edu/html/webx/index_"
29     tmp2="._html"
30     tmp=${tmp1}$blah$tmp2
31     echo "  ${tmp} being processed"
32     if [ -e "${tmp}" ]; then
33         echo "    ${tmp} already exists"
34         cp index.html /srv/www/vhosts/hrs-staff.isr.umich.edu/html/webx/index.html
35     else
36         mv -v /srv/www/vhosts/hrs-staff.isr.umich.edu/html/webx/index.html
          /srv/www/vhosts/hrs-staff.isr.umich.edu/html/webx/index_$blah._html
37         echo "    old index.html backed up"
38         cp index.html /srv/www/vhosts/hrs-staff.isr.umich.edu/html/webx/index.html;
39     fi
40
41     echo "  index.html copied"
42     echo " "
```