



Tracking Users at a NOAA Data Center



Methods

Examples

Govt. Limitations

Future Plans



Methods at NGDC



Analog

The most popular logfile analyser in the world

- Monthly Harvesting
- Adhoc Processing
- Quarterly Reporting
- Annual Summaries

[FTPServers/](#)

[Test.WebServers/](#)

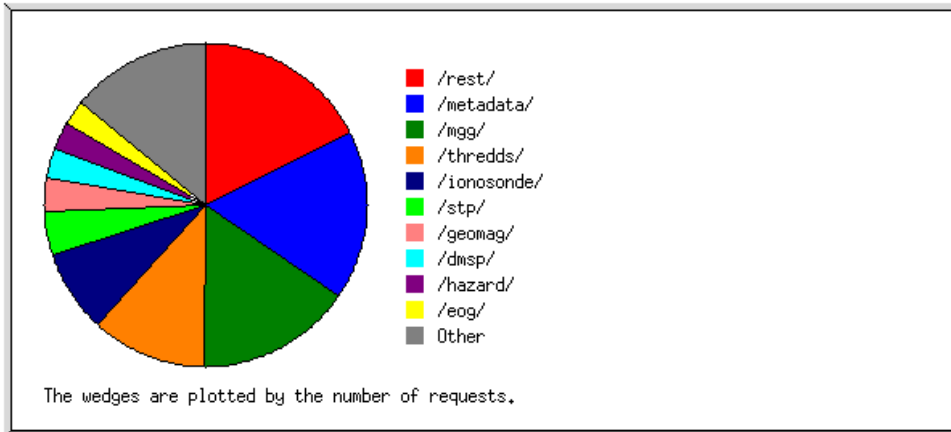
[WebServers/](#)

[WebServersSpecialReports/](#)

<http://www.analog.cx/docs/reports.html>

Summary Report

day:	#reqs:	#pages:	
---	-----	-----	
Sun:	1981124:	361981:	+++++
Mon:	3527755:	423221:	+++++
Tue:	2723905:	457900:	+++++
Wed:	2586671:	421355:	+++++
Thu:	5100922:	570917:	+++++



7:	1035093:	147513:	+++++
8:	1066720:	148634:	+++++
9:	1039733:	162669:	+++++
10:	1221180:	182337:	+++++



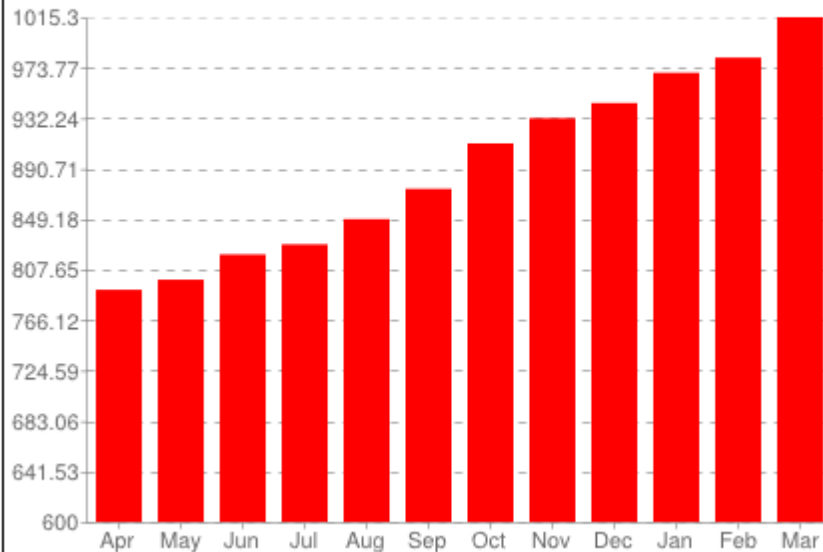
Adjustments to “Data Delivered”



Archive Size (1 Copy as of Apr. 1, 2012: 518.4 TB.) [↗](#)

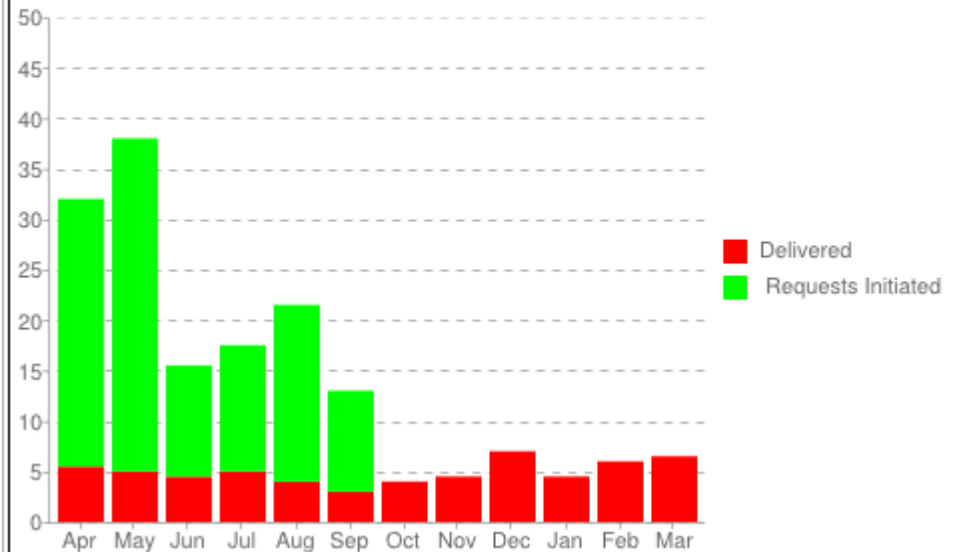
Avg. Growth Rate: 7.7 TB per Month

FY11 - FY12 Cumulative Onsite and Offsite Volume (TB)



Online Data Requested/Delivered per Month [↗](#)

FY11 - FY12 Requests Initiated/Delivered (TB) per Month





The Monthly “Rich” Report



Having some fun with grep, we discover it was indeed UCAR who hit us those days:

```
[webstat@apex vertex.ngdc.noaa.gov]$ zgrep -c ucar.edu anon-vsftpd.log-201201*
```

```
...
anon-vsftpd.log-20120116.gz:4136
anon-vsftpd.log-20120117.gz:4141
anon-vsftpd.log-20120118.gz:3897
anon-vsftpd.log-20120119.gz:1560265
anon-vsftpd.log-20120120.gz:2476789
anon-vsftpd.log-20120121.gz:614
anon-vsftpd.log-20120122.gz:718
```

Looks like they were crawling our ionosonde files:

```
[webstat@apex vertex.ngdc.noaa.gov]$ zgrep ucar.edu anon-vsftpd.log-20120119* |head
```

```
Wed Jan 18 14:40:03 2012 1 ursa.cosmic.ucar.edu 2404 /ionosonde/mids02/AS00Q/individual/2002/060/scaled/
AS00Q_2002060000005.SAO b _ o a -wget@ ftp 0 * c
Wed Jan 18 14:40:03 2012 1 ursa.cosmic.ucar.edu 958 /ionosonde/mids02/AS00Q/individual/2002/060/scaled/
AS00Q_2002060000505.SAO b _ o a -wget@ ftp 0 * c
Wed Jan 18 14:40:03 2012 1 ursa.cosmic.ucar.edu 4329 /ionosonde/mids02/AS00Q/individual/2002/060/scaled/
AS00Q_2002060001005.SAO b _ o a -wget@ ftp 0 * c
Wed Jan 18 14:40:03 2012 1 ursa.cosmic.ucar.edu 2630 /ionosonde/mids02/AS00Q/individual/2002/060/scaled/
AS00Q_2002060001505.SAO b _ o a -wget@ ftp 0 * c
```

And those ionosonde hits came from a single machine and accounted for 98% of all the hits from UCAR last month:

```
4 [webstat@apex vertex.ngdc.noaa.gov]$ zgrep ucar.edu anon-vsftpd.log-201201{19,20}* |grep ionosonde |wc -l
1 4037039
1 [webstat@apex vertex.ngdc.noaa.gov]$ zgrep ucar.edu anon-vsftpd.log-201201{19,20}* |grep ursa.cosmic.ucar.edu |wc
-1
4037039
```



Quarterly Reporting Example



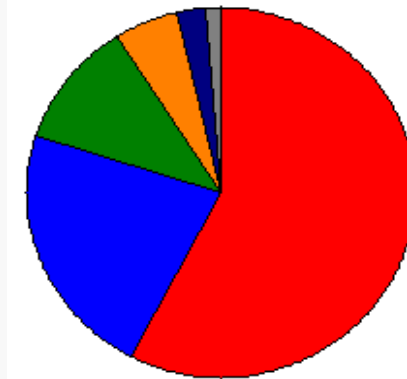
Web Hits

11,555,009 (~128K / day)
• 56% increase from Q1



Hosts Served

666,958
• 73% increase from Q1



mgg	57.58%
hazard	22.31%
geomag	16.44%
geomagmodels	
arcgis	03.02%
Other	00.65%

wedges plotted by # of requests
/geomagmodels/struts/calcDeclination = #1 (3% of reqs)

Top non-image* HTTP Hits

- /geomagmodels/struts/calcDeclination*
- /hazard/cgi-bin/*
- /dem*
- /tsunami_eq IDB*

* etopo1 WMS hits counted as images = #1 arcgis hit/#3 overall request

FTP Hits

159,091 (~11K / day)

FTP Hosts Served

5,752

FTP Volume

Downloaded

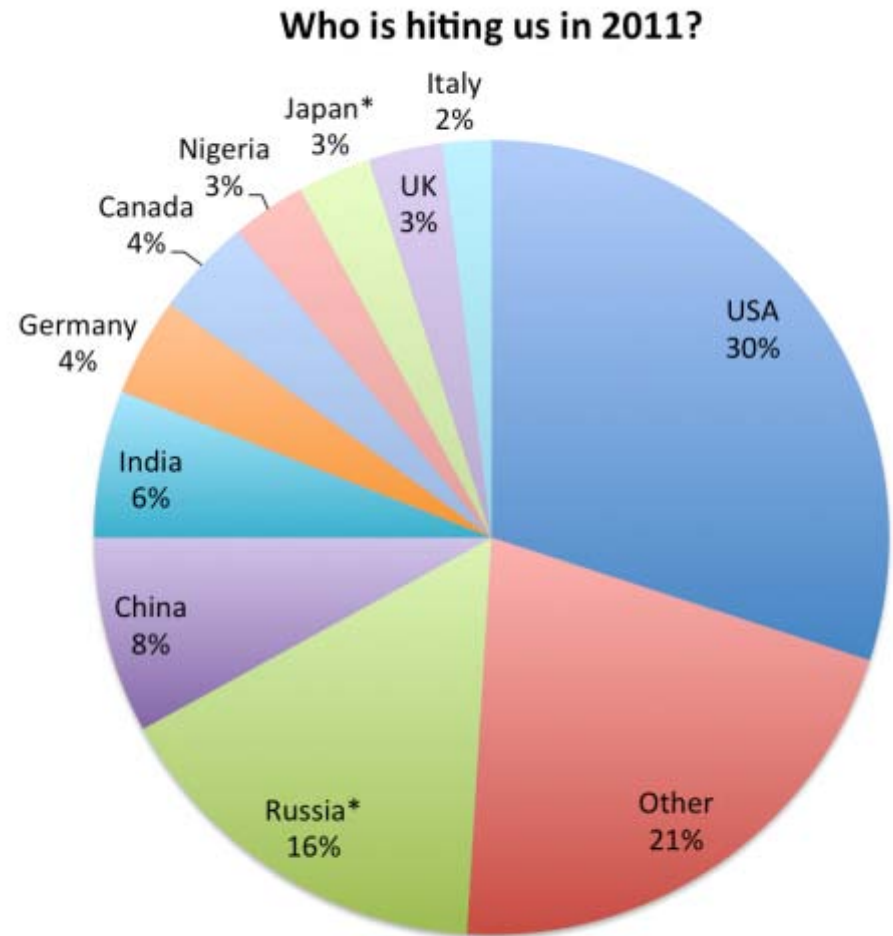
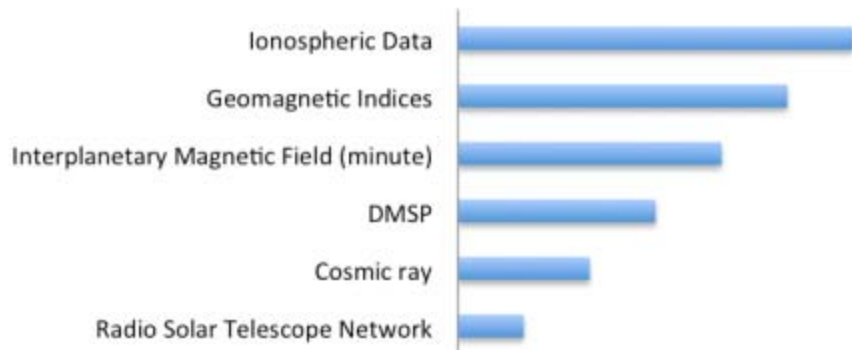
1.18 terabytes
(~14.17 gigabytes / day)

Annual Reporting Example



- 294,337 visits (browser/user only)
- 14,658 unique visitors
- 9.27 pages/visit
- 6:45 avg. duration

Most Accessed out of 28 Data Sets:



* SPIDR NODES



Alternatives: Google Analytics?

Advisement from within NOAA:

No NOAA sites are currently approved to use Google Analytics. Which is not the same as "is anyone using Google Analytics." But no NOAA sites have gone through the approval process.

There is another issue having to do specifically with NOAA: the DOC policy on social media and third party sites requires social media accounts to be tied to official government accounts. Seems reasonable. Except that **the UMS team blocked all [noaa.gov](https://www.noaa.gov) accounts from using Google Analytics**, Feedburner, Google Webmaster Tools, etc. So until this is resolved, you can't use Google Analytics.

As for the Tier 1, Tier 2, etc., yes, you do have to do a privacy review of any site that you intend to use with Google Analytics, and you **must adjust your Privacy Policy to state that you are exporting your traffic to a third party for analysis**. Another big problem is the *requirement* in the OMB policy that **visitors be given the opportunity to opt out**. Many government sites have decided this means telling visitors that they should turn off JavaScript if they don't want to be tracked -- but this leads to some rather difficult technical questions. The chief one being, of course: does your site work with JavaScript turned off?

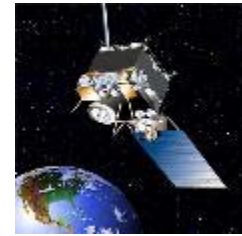
There are lots of worms involved with Google Analytics. If and when the CIO Council removes the restrictions on Webmaster Tools, Feedburner, Google Analytics, etc., I suspect we'll spend more time thinking about how to use Google Analytics correctly and within policy. Right now, / we're stuck.



Future: Convergence with CLASS



NESDIS





CLASS Metrics Reporting

Activity by Class Node		
CLASS NODE	TOTAL LINE ITEMS	TOTAL VOLUME BYTES
NCDC	21,177	4,441,434,813,112
NGDC	79,012	2,512,143,814,889
NSOF	145,702	13,840,378,967,327

Activity by Date								
ACTIVITY DATE	NCDC LINE ITEMS	NCDC VOLUME	NGDC LINE ITEMS	NGDC VOLUME	NSOF LINE ITEMS	NSOF VOLUME	TOTAL LINE ITEMS	TOTAL VOLUME BYTES
2012/04/03	9,693	2,124,458,997,043	31,451	1,066,507,517,825	54,604	6,131,740,090,471	95,748	9,322,706,605,339
2012/04/04	9,199	1,764,545,571,832	35,965	1,131,267,842,007	58,116	5,627,435,251,938	103,280	8,523,248,665,777
2012/04/05	2,285	552,430,244,237	11,596	314,368,455,057	32,982	2,081,203,624,918	46,863	2,948,002,324,212

Produced on 4/6/2012 @ 4:42:31

- >> User Profile
- >> User Preferences
- Advanced Options
- >> Job History
- Release Info
- >> Version 5.5.1
- March 28, 2012
- Other Links
- >> CLASS Home
- >> NODC
- >> NCDC
- >> NGDC

Include the Following Order Types:

- Large
- NEAAT
- NES
- Online

Include the Following Delivery Methods:

- FTP
- FTP Push
- NES

Include the Following Item Class Nodes:

- NODC
- NGDC
- NSOF

Include the Following Process Stages:

- Delivery
- Staging
- Submission

Data Center Requirements for Future Implementation:

- Track Orders by Domains
- Track Orders by Organizations.



For the Future: Like NOS, Go Social!



- □ Twitter: ~ 36,500 followers

twitter



- □ Facebook: ~ 6,730 'likes'



- □ YouTube: ~ 715 subscribers, ~ 88,000 video views



- □ Flickr: ~ 70,000 image views

flickr

- □ Stumbleupon now accounts for +25% of traffic to the NOS website



- □ Between 2010 and 2011, the NOS website had a 25% increase in visits

- □ Since 2008, site visitation has increased by more than 65%



Questions?



dan.kowal@noaa.gov
303-497-6118