

U.S. Department of Commerce
National Ocean and Atmospheric Administration
National Weather Service
National Centers for Environmental Prediction

Technical Note

Ocean Modeling Branch and the Web
Robert W. Grumbine

This is an unreviewed manuscript, primarily intended for informal exchange of information among the NCEP staff members.

OMB contribution 187

Abstract

The Ocean Modeling Branch (OMB) of the Environmental Modeling Center has been on the World Wide Web since October 1995. All Branch operational products are now available on the web site. Since its inception, OMB usage has expanded and the web site has many more users who are making serious use of products available on the WWW and by ftp. This note discusses the history, present status, and future plans for the OMB web site.

1 Introduction

OMB now has four years experience on the web. In that time, we have gone from having a few products available and a few visits per day, to having all operational products and most experimental products available, and receiving thousands of visits per day. Within the NWS, and OMB, the web and ftp distribution of information has shifted from a relatively low to a high priority.

OMB experience has shown that the web is an effective way of providing early prototypes of products to NWS field offices in time for there to be feedback from the field before final implementation. Making products available in this way has brought a large number of outside users to NWS products, who would never have known we were even producing marine products.

Because the web, and our use of it, is changing rapidly, any static document will become dated. Nonetheless, it is useful, to have a snapshot of what our status is at a given time, in a format which can be used away from the web itself. This will be the first of a periodically-updated series each of which will describe the OMB's evolving use of the web. It will also contain annual summaries of our web and ftp usage statistics.

The following sections are: Policy considerations, OMB standards and practices, statistics for fiscal years 1998 and 1999, thoughts towards the future, and conclusions.

2 Policy Considerations

The main policy considerations from NWS headquarters are to:

- 1) Provide access, and restricted access, commensurate with international agreements on data distribution.
- 2) Provide equal access to all commercial entities.
- 3) Clearly inform regarding the priority of web services within the NWS.

Point 1 is addressed in a memo dated 31 May 1996. In brief, it means that we cannot redistribute data from foreign sources that we obtained under a restricted access ourselves. We may, however, display data that we have significantly reprocessed.

Point 2 is a continuing concern of commercial weather services. There are also concerns regarding NWS web pages competing with commercial entities. The OMB, for example, is referred to in the Commercial Weather Services Association legislative agenda: http://weather-industry.org/legislative/Legislative_Agenda.html NWS decision is that model output and graphics may be made available (indeed, should be made available), but it must be generic

output, generic graphics, and cannot be a model or graphic which is tailored to a specific industry or corporation. Letting users produce customized figures, such as the sea ice page, where users may select areas to magnify, with a general tool is acceptable.

On the third item, a Jack Kelly memorandum dated 4 January 1999, requires that NWS servers link to a NWS disclaimer. OMB's web server falls under the 'non-operational' category, so the appropriate disclaimer is <http://www.nws.noaa.gov/disclaimer1.html>. This requirement applies only to the server itself, i.e., the main page on the server. The OMB standard is to link main pages for each product set to the disclaimer as well.

It is NWS policy that Internet information distribution is subject to the same supervisory and organization approval mechanisms and procedures as distribution by other methods. Because of this, approval for additions to web pages should be obtained through webmaster Robert Grumbine (designated reviewer for Internet products by Branch Chief D. B. Rao), the Branch Chief, or as part of the standard approval process for other documents such as TPB's, tech notes, etc.

Regarding what can be made available on the net, NWS regulations state:

" a. No special products should be developed solely for the Internet. Products acceptable for display on the Internet include:

- NWS products that can now or will in the future (see item b. below) be obtained from operational NWS dissemination services (e.g., FOS, NWWS, NOAAPORT, etc.)
- Representations of individual NWS products to which no value has been added. Embellishments such as the use of color, animation, or multiple image display, which are made possible through Internet technology, are not considered adding value to base products.

b. In general, products labeled "experimental", "non-operational" or "prototype" may be displayed on, or accessed via, the Internet if permission is granted at the Office Director, Region Director, or Center Director level with notification to the Industrial Meteorology Staff (W/IM)

"

Products that we call 'experimental' are actually under part 'a'. Part b is what we're calling 'developmental' and not making generally available. Our use of 'developmental' is what is typically considered 'Intranet' rather than Internet.

3 Practice: OMB web page standards and tools

For the most part, this section will describe standards we have implemented. Please verify that your pages comply with the standards. A later subsection will note what tools we have available for constructing web pages and carrying out the required practices.

3.1 Standards and Practices

OMB web standards and practices are more restrictive than the NWS limits. Where headquarters has made a requirement, it will be noted as the source. Otherwise, the restriction is a requirement OMB has agreed to ourselves and is subject to reconsideration. There are four categories of items listed: "required", "recommended", "deprecated", and "forbidden". Practices not listed can be considered optional – do them if you like, and we'll decide if this

is a good idea after you've used them for a while. If users tell us it is wonderful, we'll copy it through the Branch, if they tell us they hate it

Required:

NWS: contact name and address

NWS: modification date

NWS: If your page has a pointer to a private sector weather products home page, you should have a link to the IM home page <http://www.nws.noaa.gov/im/>

Time specified as UTC rather than Z

Graphics must have indication that this is a NOAA/NWS product

Each page must link to at least one higher level page. For example, the sea ice analysis page must link to at least the main sea ice page. The main sea ice page must link to at least the main Branch page. (NWS requires main site page to do so, OMB requires all pages)

Branch products such as waves, ice, ... should link their main pages to the NWS disclaimer. (NWS requires main page to do so).

Test pages with lynx (a text-only browser available on any linux box)

Test pages with weblint (weblint page.html)

ALT tags for inlined graphics (see 'tools' subsection for how)

Logos for models/product groups/ etc. should be placed in the omb/images directory.

Recommended

White background

Graphics should be in .gif

DOC, NOAA, NWS, NCEP logos on main page of product suites

'Keywords' in the page headers to improve indexing by search engines

Consider linking product main page to mailing list sign up page

NWS headquarters has made tools available for web developers, at <http://tgsv5.nws.noaa.gov/project/developer.html>

This includes things of interest to any web page developer, as well as some only of concern to webmasters.

For presentation, the NWS guideline notes: "HTML documents should, for the most part, use HTML 2.0 syntax. Non-standard HTML features, such as HTML+, HTML3.0, HTML 3.2, and JAVA should be used with care so that data is not denied to the public. Pages should be constructed without reliance on graphics to support text-only or low access speed clients. For example, an image map should be accompanied with an alternative means of accessing the links embodied in the image map."

Deprecated

graphics in .jpg (jpeg) format. For our products, jpeg results in larger graphics. jpeg is also a lossy format (it doesn't exactly reproduce the original) it is undesirable on both counts. JAVA is usable by only certain classes of user, and making similar information to others can be difficult.

Frames. If not used very well, the result can be very bad, to the point of violating our desire

to make information available to the widest possible audience. Irrespective of use, frames have a tendency to crash one of the two most popular web browsers.

Postscript files on the web page: These are only viewable if the user has an appropriate postscript viewer. Linux boxes do, but many users don't have linux. Nonetheless, postscript is good for ftp distribution.

Forbidden:

Graphic-only navigation.

Graphics in TIFF format. TIFF is a non-lossy format but its files are huge, typically an order of magnitude larger than gif. It also has the drawback that it is not one of the formats which browsers can be assumed to display.

Graphics of hit counts.

3.2 Tools

A separate document will describe how to write HTML, including describing tools for doing so, and note what features are in which markup standard.

3.2.1 Logos

Logos and other small graphics can be attractive additions to web pages. There is a specially-designated location for such things: /data/WWW/omb/images. There are a couple of reasons for using a single location for such graphics. First, it is handy to have a location from which we can share images – should anyone want buttons or numbers there is a set already available there. There are already DOC, NOAA, NWS, and NCEP logos in that directory, and they have been carefully selected so as to be the same size and graphically comparable. Be careful that these little graphics don't become a large portion of the load time. Also, if you have an image 'inlined', i.e., you use `` you must add one more element to the reference: ``. The "alt" is text which is displayed when the user either has a browser that does not display graphics, or has shut off automatic graphic downloading.

There is currently one exception to the policy regarding logo placement. That is 'favicon.ico'. This entity is used by MS Internet Explorer when a user bookmarks a page as one of his favorites. When displaying his bookmark list, the icon is shown next to the name. There is now a link to the OMB logo as 'favicon.ico' in directories from which I've seen users getting this error. I recommend that if you have this file in your web directory, you should create an 80x80 logo in gif format and give it this name. xpaint and xfig, available on workstations and linux boxes, are useful for making this graphic.

3.2.2 Animations

The easiest way to construct an animation is 'whirlgif'. It is available in /usr/local/bin on polar, and in /nfsuser/g01/marine/local/bin on the IBM SP. This requires that the input

images be in gif format. The output is just a different version of the gif standard and is automatically viewable in graphic web browsers. The file size is the sum of the sizes of the input images. A different, more compact, format is mpeg. gifs may still be used as input, but the software produces an mpeg output. This is not automatically viewable on web browsers, but mpeg animations may be twice or more times as compact as gif animations, as mpeg examines consecutive images and only saves the portions which change. Software for doing this is in testing.

3.2.3 Making .gif

Most of our graphic packages, c.f. grads, ncargraphics, IDL, produce postscript output, rather than gif. To convert to gif, use pstogif. You can also use pstogif2 to do a conversion with automatic cropping. The cropping will trim to the boundaries of all visible text, dropping the constant color background. On polar you need to have /usr/local/bin/ and /usr/local/bin/netpbm in your PATH. On the SP, now the preferred location for image conversion, you need /nfsuser/g01/marine/local/bin and /nfsuser/g01/marine/local/bin/netpbm

3.2.4 Restricted access distribution

We can make materials available to a selective set of users by the web, ftp, and mail. For the web, things placed in the 'develop' directory are limited by default to machines on the floor, specifically the 192 network. Specific sites may be added to this approved list. This is control by machine identity, and does not work when the access will be made from PPP connections, AOL for example. For that case, we have a password protection ability. This requires users to go to a different server, <http://polar.wwb.noaa.gov:8000/> where the 8000 specifies the password side.

For ftp, we've established a directory for restricted distribution. This is /home/ftp/pub/tempor. You must put your files there as yourself (i.e., you can't use anonymous ftp to upload the files). Once there, the file can be retrieved anonymously. Warn users that they will not be able to do an 'ls' or 'dir'. They must (and can!) simply say 'get filename' after they've entered cd pub/tempor. As the name suggests, this isn't a directory in which to leave things permanently. If you have a file that needs to be available with limited access for an extended period, say greater than 3 months, see the webmaster about other methods.

We can also establish mail accounts and handle automated responses to messages. This is the method being used to manage model source distribution, for example. This permits you to track requests automatically.

3.2.5 Source Distribution

We handle source code distribution through automatic response to mail requests and restricted access ftp directories. When you have source to distribute this way you need to:

1. Document the file(s) and instructions to open the distribution archive.
2. Package the files for distribution including constructing a tar file and constructing a copy of the tar file that is gzip'ed.

3. Determine a unique subject line that your users will use for the mail request.
4. See the webmaster to implement the mechanism.

We can also distribute data this way, but this hasn't been a need yet.

3.2.6 Mailing List

We have a mailing list for occasional announcements. I've been sending a message every few months. The sign up page is at <http://polar.wwb.noaa.gov/omb/mailing.html> If you don't already have a link to this list from your main product page(s) you should consider doing so. If you have material you'd like sent out to the mailing list, please inform me. We now have about 300 people signed up.

3.2.7 Weblint – html standard conformance

Many of the tools for web page writing will do nonstandard things. Some of our default practices have skipped things which should be done for complete standard compliance, but which we can generally get away with. Since we do want be accessible to the widest audience possible, we need an html checker. weblint is one such. The version currently on polar will flag some things which are now acceptable. But it does catch a lot of errors that are easily avoided even so. To check your page, type 'weblint file.html' (you need /usr/local/bin/ to be in your PATH). If the page is long, you'll want to redirect output to a file in your home directory. There is a newer version of weblint in testing. Retest your pages after every change you make.

3.2.8 Tools for web page construction/editing

Netscape Composer and Applix have html editors. Word processors can export to html, and there is a LaTeX to html converter. I've tested none of these to any great degree as yet. Instructions for how to write html (including how to use these tools) will be a separate document.

3.2.9 Keyword usage

Keywords in your web page enable search engines to your pages more effectively in response to user requests. Web page editors let you specify keywords and the OMB Welcome page has an example you can follow manually. More in the 'how to write html' document.

3.2.10 Web page templates

There will be web page templates/prototypes available in `web/templates`. These will be suitable for general purpose Branch web pages. More examples will be in 'how to write html'.

3.2.11 CGI

CGI scripts and programs can be used to let users customize the information they retrieve from OMB, or to select particular data set(s) from many possibilities. As such, CGI's are very good things. Nevertheless, they have two drawbacks we need to consider: CGI's are, in principle, security risks since it is possible for a cracker to subvert a script or program so as to do malicious things. It is easy for them to do so if you don't use care in writing your CGI. Also they put an extra load on the server. Because of the load issue, our current requirement is that CGI's finish in less than 5 seconds on the branch web server. This limit may be modified if the CGI is used heavily, or especially lightly. Because of the security question, it is required that the CGI's be tested on lenox1, and only be installed to polar by the webmaster. A separate document regarding CGI's is in preparation, along with examples.

4 Statistics FY 1998 and 1999

4.1 Web

Figure 1 presents the monthly visitation statistics from October 1997 through September 1999 exclusive of errors and WWB usage. In that time, typical usage has increased from about 35,000 per month to over 180,000 per month. Much of the increase was in two pulses: in early 1998, and again from June through September 1999. The latter period corresponds to the increasing availability and publicity of the NOAA WaveWatch III model. The former seems to have been spontaneous. Also spontaneous is the spike in visitation in November 1998.

As a rough extrapolation, we could take a factor of 5 increase in visitation rate per two years and guess something like 1 million hits per month, 40,000 per weekday, towards the end of FY 2001. The highest usage yet experienced was 16,000 hits on a single day, the day an article citing the branch's page was published in USA Today (June 1998). That load posed no difficulty for the web server (a pentium 166 with 10baseT, running linux, with the CERN server software), so we would expect no difficulty for another year according to our extrapolation. The branch is staying ahead of load by moving to a new, faster, server in June, 2000.

The majority of OMB page visitation is from U.S. sites, with over 1 million hits lead by network service providers. For perspective, a heavily visited page on the net can expect a million hits per day. The record seems to be 30 million hits in a single day for one of the Mars Sojourner pages. We can get an estimate of which countries access the OMB web site from sites which provide a verbal name. Many sites provide only the machine number, so this listing is incomplete and where numbers of hits are given, this is a lower bound. Usage from within the WWB itself (including OMB, MPC, CPC, and other NWS/NESDIS usage) was approximately an additional 150k hits; that is, about 10% of all traffic was inside our building.

Top level domains visited from:

Andorra	United Arab Emirates	Antigua and Barbuda
Anguilla	Argentina (6.5k)	Austria (1.4k)
Australia (32.3k)	Aruba	Bosnia-Herzegovina
Barbados	Belguim (5.8k)	Bulgaria
Bahrain	Bermuda	Brunei Darussalam
Bolivia	Brazil (40.6k)	Bahamas
Botswana	Canada (23.3k)	Switzerland (4.1k)
Ivory Coast	Cook Islands	Chile
China	Colombia	Costa Rica
Cyprus	Czech Republic	Germany (19.3k)
Denmark (4.0K)	Dominica	Dominican Republic (6.4k)
Ecuador (1.1k)	Estonia	Egypt
Spain (20.8k)	Finland (1.5k)	Fiji
Micronesia	Faroe Islands	France (99.7k)
Great Britain (see UK)	Georgia (country!)	GF?
GG?	Gibraltar	Greenland
GP?	Greece	Guatemala
Guyana	Hong Kong	Croatia
Hungary	Indonesia	Ireland (7.0k)
Israel	India	Iran
Iceland (2.1k)	Italy (7.5k)	Jamaica
Japan (17.7k)	Kenya	Comoros
South Korea (1.5k)	Kuwait	Cayman Islands
Kazachstan	Lebanon	Liechtenstein
Lithuania	Luxembourg	Morocco
Monaco	Madagascar	Macau
Martinique (French)	Malta	Mauritius
Maldives	Mexico (3.6k)	Malaysia (2.1k)
Namibia	New Caledonia (French)	Niger
Nicaragua	Netherlands (24k)	Norway (17.9k)
NS?	Niue	New Zealand (5.5k)
Oman	Panama	Peru (5.7k)
Polynesia (French)	Papua New Guinea	Phillipines
Poland (3.0k)	Portugal (4.6k)	Paraguay
Qatar	Romania	Russia (2.0k)
Saudi Arabia	Sweden (3.7k)	Singapore (1.1k)
Slovenia	Slovak Republic	San Marino
Senegal	Soviet Union	El Salvador
Turks and Caicos Islands	Togo	Thailand
Tonga	Turkey	Trinidad and Tobago
Taiwan (28.6k)	Ukraine	Uganda
UK (102.8k)	US	Uruguay
Vatican City State	Venezuela	Virgin Islands (US)
Vietnam	Vanuatu	Yemen
Yugoslavia	South Africa (1.3k)	Zimbabwe

Top level domains not visited by:

Afghanistan	Albania	Armenia
Netherland Antilles	Angola	Antarctica
Azerbaijan	Bangladesh	Burkina Faso
Burundi	Benin	Buthan
Bouvet Island	Belarus	Belize
Cocos (Keeling) Islands	Central African Republic	Congo
Cameroon	Czechoslovakia	Cuba
Cape Verde	Christmas Island	Djibouti
Algeria	Western Sahara	Ethiopia
Falkland Islands	Gabon	Grenada
Guam	Heard and McDonald Islands	Honduras
Haiti	British Indian Ocean Territory	Iraq
Jordan	Kirgistan	Cambodia
Kiribati	Saint Kitts Nevis Anguilla	North Korea
Laos	Saint Lucia	Sri Lanka
Liberia	Lesotho	Libya
Moldavia	Northern Marina Islands	Marshall Islands
Mali	Myanmar	Mongolia
Mauritania	Montserrat	Mozambique
Norfolk Island	Nigeria	Nepal
Nauru	Neutral Zone	Pakistan
Saint Pierre and Miquelon	Pitcairn	Puerto Rico (US)
Palau	Rwanda	Solomon Islands
Seychelles	Sudan	Saint Helena
Svalbard and Jan Mayen	Sierra Leone	Suriname
Saint Tome and Principe	Syria	Swaziland
Chad	French Southern Territory	Tadjikistan
Tokelau	Turkmenistan	Tunisia
East Timor	Tuvalu	Tanzania
US Minor Outlying Islands	Uzbekistan	Saint Vincent and Grenadines
Virgin Islands (British)	Wallis and Futuna Islands	Samoa
Zambia	Zaire	

4.2 FTP

Table 1 gives the monthly ftp site usage, in terms of numbers of files remotely accessed and the megabytes of files transferred. The peaks in July-September are partly due to hurricane season, and partly due to a shift in the files we make available. Almost all the ftp traffic is for wave model output. Exclusive of this, the next is sea ice climatic information, which totalled about 500 Mb in FY 1999, rather less than 10% of a single month's tally for wave output. In 1999, we began producing and making available the NWW3 model output. This is more compact than the old global ocean model, by a factor of approximately 4. The NWW3 output is also in fewer files, so that although the numbers of accesses and bytes

transferred didn't increase markedly from September 1998 to September 1999, the amount of usable information transferred increased by a factor of 4-5.

Month	Accesses	Mb
Jan 1998	2479	4957.6
Feb 1998	5039	9013.7
Mar 1998	9739	5589.5
Apr 1998	10562	5655.1
May 1998	13379	7332.9
Jun 1998	12403	7266.6
Jul 1998	21718	17076.2
Aug 1998	27954	22299.3
Sep 1998	25384	18235.7
Oct 1998	19836	12979.7
Nov 1998	11998	7558.2
Dec 1998	14154	9335.3
Jan 1999	16648	9431.3
Feb 1999	14107	9139.4
Mar 1999	32610	30589.6
Apr 1999	22715	27269.3
May 1999	29053	26843.1
Jun 1999	29711	22360.1
Jul 1999	30016	18996.5
Aug 1999	34392	21943.5
Sep 1999	38371	25150.5

Usage by country and site follows similar patterns to web page visitation. Again, many of the countries of the world have retrieved files from the ftp server. Some countries are much more active on the ftp transfer than is suggested by the web side, Brazil and Argentina are two such. Some are rather less active, such as the UK.

One feature of the ftp usage, in contrast to web visitation, is that there is no clear drop-off in usage on the weekend. Total load is fairly constant. In FY 1999, the weekend had somewhat higher traffic than usual, with Wednesday being the low point. Taken as a steady rate, the 25 Gb transferred in September 1999 represents about 35 Mb per hour, or about 10 Kb per second. This is equivalent to two 56 Kb modems operating full-time, or about 10% of a T1 connection. Since this is average load, and one can expect the usage to be varying through the day, this is suggestive that our data transfer is being limited at least part of the time by bandwidth out of the building. The drop in mid-week may also be due to this, as this is when WWB internal activity is high.

Since the overwhelming majority of data transfer involves the wave models, it will be worthwhile to explore more efficient storage and transmission methods. Since it appears we're already being affected by bandwidth limits, we should consider further methods of either improving our bandwidth or sharing the communication load with points outside

WWB.

5 To the future

Web and ftp distribution took a major step forward in fiscal 2000, when it was made one of the primary groups for OMB task plans. Information regarding OMB's plans for the web is therefore in the OMB task plan. We are continuing to work on, expand, and improve, the OMB web pages. This includes shifting, much of the generation of web-related material to the central computing facility.

A general point regarding the future is a reflection from the past. OMB has run the Branch web and ftp distribution locally. The webmaster is in the Branch, the machine is in the branch, and control of the system is local. This has meant that when there is a problem with the page, whose resolution OMB considers a high priority, the people responsible for fixing it and the person needed to get it fixed all consider the repair a high priority. Since control is local, OMB has also been able to add features and hardware in a timely manner. Some things are now sufficiently routine that responsibility can be handed over to another group or organization, for example updating the software we use for image conversion. Many others, however, still require local control, and this is expected to continue into the future.

Most importantly, the web and ftp comprise the future of OMB product distribution, whether analysis, guidance, image, data, source codes, or documents. We can expect continued evolution in our web and ftp practices as increases in volume impose new requirements, and as new technologies become available.

6 Conclusions

OMB has a history of good experiences from our web and ftp distribution of branch products. Our users span much of the world, and are generally appreciative of what we do. Without the web, many of these people would never have used our products, or even known that we existed, without the web.

The current version of this document will be available on the web as `develop/omb/ombweb.html`. A mirror of the most recent paper version will be in `omb/papers/tn187/tn187.html`.