Web Analytics and Site Design

Lawrence I. Charters
National Ocean Service
NOAA WebShop 2007
November 14, 2007

Web Analytics and Site Design
Lawrence I. Charters
National Ocean Service
NOAA WebShop 2007
November 14, 2007

Photo of Hadrian's Wall by Lawrence I. Charters

Starting With False Assumptions

Starting With False Assumptions

Photo of Air Force Thunderbirds Falcon F-16 shortly before crash at Mountain Home Air Force Base, Idaho. The pilot had set his ground indicator at an air base with a lower altitude, and after doing a high-energy maneuver at Mountain Home AFB, found himself much closer to the ground than he planned. He had minor injuries; the $30 million aircraft was destroyed. USAF photo by Staff Sergeant Bennie J. Davis III, a man with nerves of steel.

Everything You Know Is Wrong: Commercial Web Analytics Don't Work

- **Hits don't matter - Wrong**

- **Page views are the only thing important - Wrong**

- **Visits are the only thing important - Wrong**

- **Only worry about the popular stuff - Wrong**

Commercial Web Analytics don't work for government sites.

Hits don't matter - Wrong
    Web servers log file transfers; one file transferred, one log entry.
    If the hit is a 300 page PDF of a report to Congress, it is important, yet is only one hit.
    If the hit is a high resolution aerial photo of storm damage from Hurricane Katrina, it is important, yet is only one hit.

Page views are the only thing important - Wrong
    A 300 page PDF is not considered even one page.
    Pages delivered using REST (Representational State Transfer) end in ordinal numbers; there is no way for a poor little Web analytics program to know that all ordinal numbers should be counted as pages.
    Frequently the content of the entry -- the image, the sound, the animation, the video -- is the most important element; the page is merely a delivery framework.

Visits are the only important thing - Wrong
    A visit is a derived metric, only indirectly based on logs: a period of activity from a visiting IP address is considered a visit; it is considered to end after an arbitrarily defined period of inactivity.
    This concept does not work in an era of firewalls and Web proxies; all 25 million members of AOL appear as only 1,300 IP addresses; all 5,000 NOAA employees in Silver Spring appear as only a few dozen external IP numbers.

Only worry about the popular stuff - Wrong
    Old data has value. It is critical in charting global climate change, ecosystem change, biological life cycles and virtually everything else NOAA does.
    The public has paid for our data, popular and unpopular.
    The Weather Service divides the country into 770,000 grid points for making forecasts. These grid points represent the entire country, not just the popular parts.
    With the exception of the 1890 Census (destroyed in a fire in the Commerce Building), the government does keep records on people (and animals, and activities, and environmental phenomena) no longer "current;" that's part of the job.

Webmasters Have Different Needs

- Program managers want to show results

- Webmasters need to know things work

- These are not the same needs

Webmasters have different needs

- Program managers want to show results
- Webmasters need to know things work
- These are not the same needs.

Discovering Causes of Failure

Discovering causes of failure

Photo of St. Andrews Cathedral, St. Andrews, Scotland, by Lawrence I. Charters. The cathedral, on a bluff over the North Sea, was destroyed by gales in 1270 and 1409.

Different Kinds of Failures

Recent Errors /Day ?	Errors ?	Latest Hit ?	Description ?
2,521.6	864,059	09/30/07	/includes/get_pred.css Local refer problem: /get_predictions.shtml

Local refer problem: Webmaster goofed

178.7	74,235	09/30/07	/data_options.shtml External refer problem: http://www.ptguide.com/weather/index.html http://atmos.uprm.edu/othersources.html

External refer problem: Webmaster may have goofed (moved something)

Different Kinds of Failures

When people get a 404 (File not found) it could be because the Webmaster goofed.

- Local referral problem: you have a link to a file that doesn't exist.
    - You never created the file.
    - You changed the path.
    - You misspelled the file name or path.
    - You redesigned the page in a GUI editor and left invisible references in the code to old files.
- External referral problem: people are linking to things that don't exist.
    - You redesigned the page.
    - You moved something.
    - The external site misspelled something in their link.

Different Kinds of Failures (cont.)

0.4	12	09/24/07	/_vti_bin/shtml.exe/_vti_rpc Possible FrontPage attack

External problem: robot attacks

0.0	1	09/14/07	/postnuke/xmlrpc.php No information available
0.0	1	09/14/07	/adxmlrpc.php No information available
0.0	1	09/14/07	/phpAdsNew/adxmlrpc.php No information available
0.0	1	09/14/07	/tiki/tiki-xmlrpc_services.php No information available
			/xmlrpc.php

"No information" but these are robot attacks, too.

Different kinds of failures:

   - Robots and worms are attacking your site, looking for vulnerabilities.

Different Kinds of Failures (cont.)

| 49.0 | 6,687 | 09/30/07 | /Barbara's+Data/C/Images/trans-top-right.gif
No information available |

Some failures even identify the responsible party: in this case Barbara, who has a Windows computer, linked to a file on her C:\ drive and not on the Web server.

| 37.1 | 21,719 | 09/30/07 | /robots.txt
Robots exclusion rules |

In this case, simply placing a robots.txt file at the root will make tens of thousands of errors go away.

Different kinds of failures (cont.)

- Some failures are obvious: they refer to files on the developer's hard drive, and may even identify the developer by name.
- Some failures are easy to fix: simply place the file on the server.

Structure Sites Based on Themes

Structure sites based on themes

      - Don't structure site after organization; you are only one reorganization away from destroying a site.
      - Don't structure site based on Webmaster's convenience; the site is for visitors, not the Webmaster's convenience.

Photo of Pigeon Point Lighthouse by Lawrence I. Charters

Form Follows Function

Directory ?	Hits ?	Sep 2007
/siteart/	2,787,370	356,489
/styles/	1,520,787	218,624
/	534,008	68,844
/foundations/	461,294	57,797
/events/	450,440	64,640
/transformations/	445,454	76,121
/magazine/	378,028	75,658
/breakthroughs/	321,520	39,298
/visions/	275,661	35,116
/scripts/	272,200	33,193
/new_york_charts/	270,548	31,668
/activities/	235,808	52,662
/historymakers/	234,310	24,819
/datasets/	217,740	29,860
/distance_tools/	187,432	24,347
/forfun/	152,097	9,526
/edufun/	148,381	18,501
/monitor/	135,085	19,924
/survey_marks/	115,588	14,895
/about/	104,893	7,403

NOAA CELEBRATES 200 YEARS of SCIENCE, SERVICE, and STE

Home Foundations Transformations Visions Top Tens Colle

About the Celebration

Top Tens
- History Makers
- The Breakthroughs
- Historic Events
- Foundation Data Sets

Feature Stories

Collections

200-Year Timeline

NOAA Historical Resources

For Kids and Educators

For Fun

Search [   ]
○ This site ○ NOAA

Site Map

Acknowledgments

Foundations: Managing Legacy Climate Data in the 20th

FOUNDATIONS

From our beginnings in 1807 to our recognition in 1970, NOAA h grown with the nation. In *Foundations*, NOAA offices and progra trace their origins and relive their **histories**, telling the stories people, ideas, events, and breakthroughs that are the foundatio upon which we stand today... **more**

TRANSFORMATIONS

The last three decades have been a period of great change with NOAA. In *Transformations*, we tell the stories of recent changes our missions, technology, and ways of doing business. These st reveal the NOAA of **today**: an organization that continues to ev to meet the needs of the American people...**more**

Form Follows Function

- Structure site (directories) thematically, not organizationally, not based on Webmaster's convenience
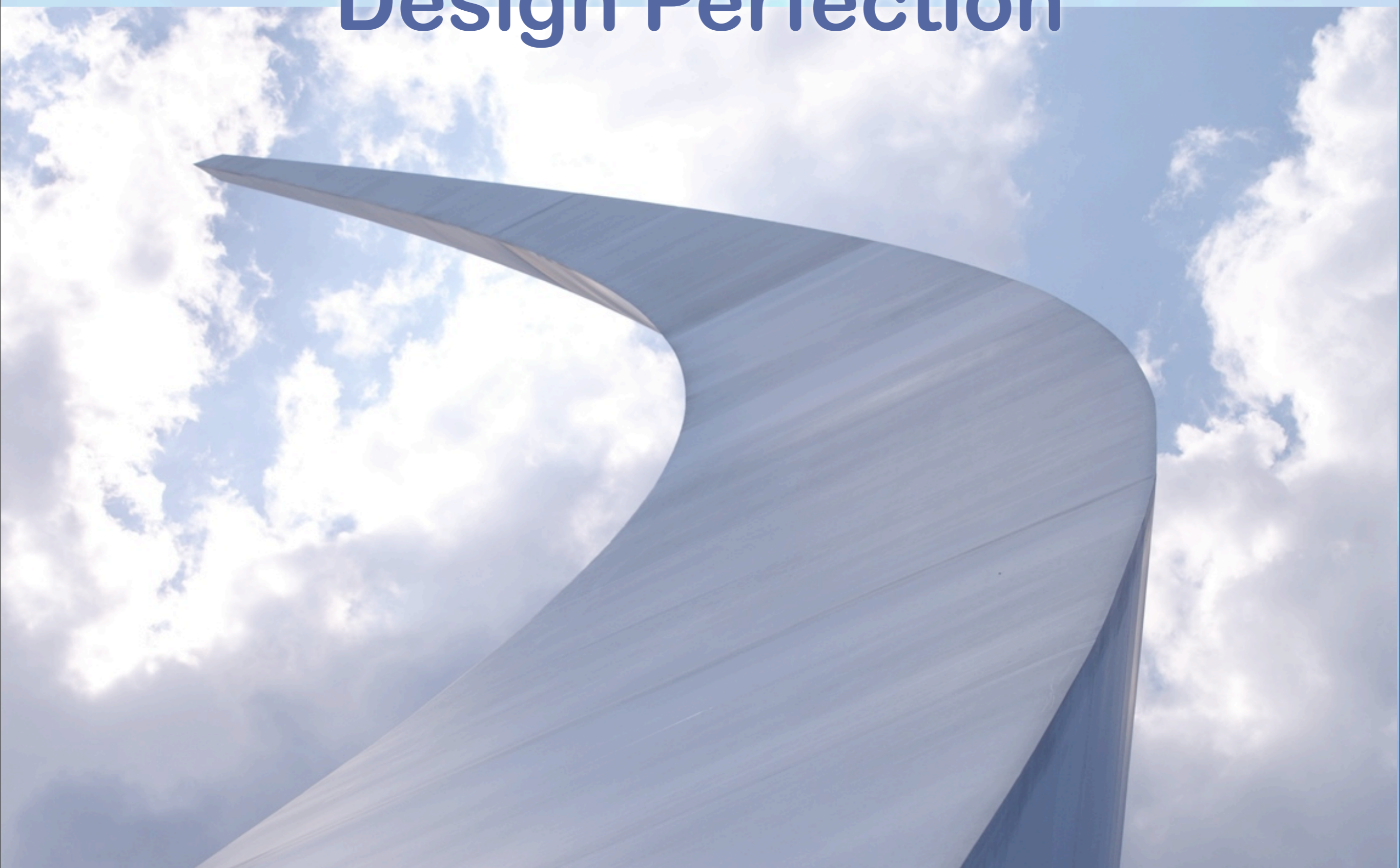- Thematic structure also leads to more accurate statistics

Design Perfection

Design Perfection

Photo of the Air Force Memorial by Lawrence I. Charters

Summary

- **Structure navigation, directories thematically**

- **Ignore organizational or political divisions**

- **Place all related materials in the same directory; do not divide by file type (without a really good reason)**

- **Get rid of silly errors; they mask structural problems with noise**

Goal

Goal:

- Make Web sites as useful and functional as libraries.
- Libraries are some of the oldest institutions on the planet.
- They outlast reorganizations, governments and civilizations.