Local Access Statistics for Federal Documents: Tracking Web Page and Online Catalog Usage
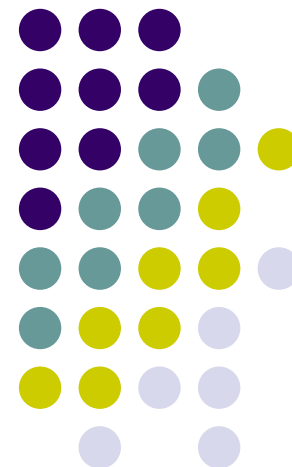
Federal Depository Library Conference
Washington, DC
Wednesday, October 20, 2004

Chris Brown
University of Denver
cbrown@du.edu

Susan Xue
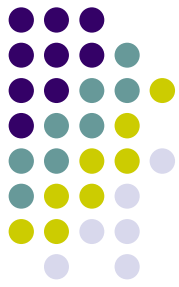University of Colorado at Boulder
susan.xue@colroado.edu

Statistics for Online Government Document Access through the OPAC : One Year of Statistics

Federal Depository Library Conference
Washington, DC
Chris Brown
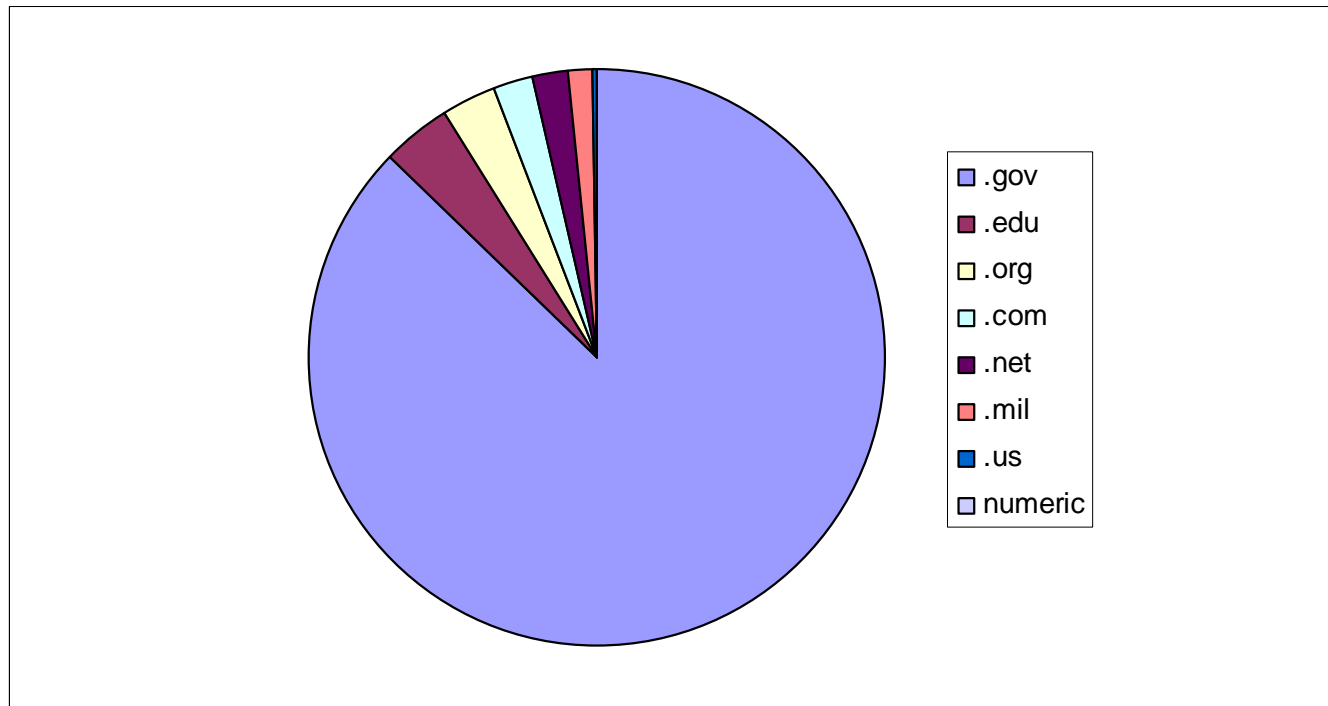University of Denver
cbrown@du.edu

The Challenge

The Challenge: Domains

- Distribution of URLs: .gov, etc.

Legend:
- .gov
- .edu
- .org
- .com
- .net
- .mil
- .us
- numeric

The Challenge: Number of Servers

- GPO counted 1,416 unique servers:
http://ww1.access.gpo.gov/gpoaccess/fdlp/tools/domains.html
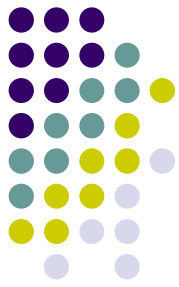
- University of Denver counted 1,398 in local OPAC

- Challenge: tracking URLs for over 1,400 unique server domains

Aggressive URL Adding

- Making of America
- Century of Lawmaking
- NASA Technical Reports
- NASA ICASE Reports
- URLs for older documents in GPO Access but not in GPO catalog records
- ERIC Digests
- Historical Publications of the United States Commission on Civil Rights (U. Maryland Law)
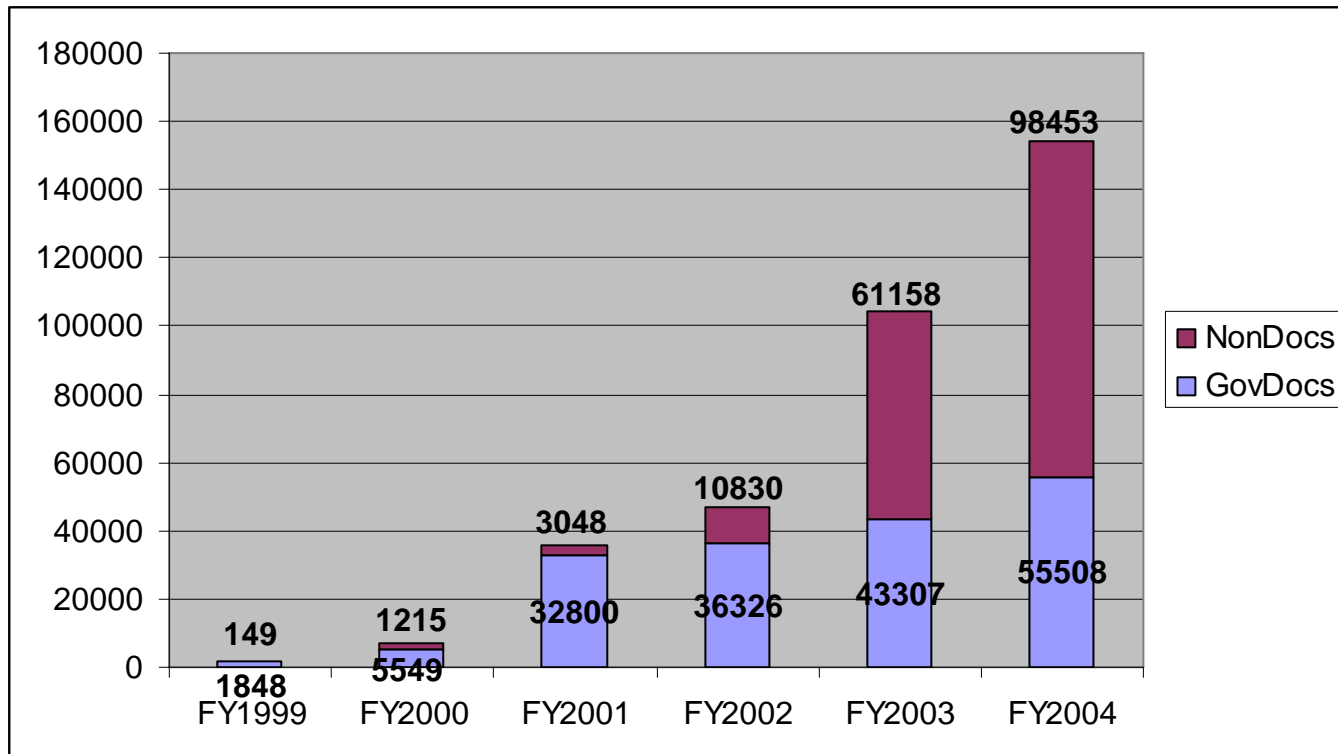
Aggressive Adding Worth it?

- Was all this aggressive URL adding worth the effort?

- Needed a way to measure the results

University of Denver Profile

- 70% selective depository
- Depository since 1909
- 80,187 URLs for govdocs, contained in 55,508 bibliographic records

URLs in DU's OPAC

The Process

Tracking System

http://library.du.edu/findit/peak/redirect.cfm?LinkURL=

We know every time a user access a government
document through our OPAC.

More on this in last year's presentation.

Last year's presentation is available at:

http://www.access.gpo.gov/su_docs/fdlp/pubs/proceedings/03prot.html

How it works

t	245	1	0	Concept mapping|h[microform] :|b
				a graphical system for understanding the relationship between concepts : an ERIC digest /|cby Eric Plotnick.
u	246	3	0	Graphical system for understanding the relationship between concepts
p	260			[Syracuse, NY] :|bClearinghouse on Information & Technology,|c[1997]
r	300			1 v.
s	490	1		ERIC digest ;|vEDO-IR-97-05
n	500			Distributed to depository libraries in microfiche.
n	500			Shipping list no.: 98-0316-M.
n	533			Microfiche.|b[Washington, D.C.] :|cSupt. of Docs., U.S. G.P.O.,|d[1997]|e1 microfiche.
n	538			Access online version: http://www.ericfacility.net/ericdigests/ed407938.html
d	650		0	Computer graphics.
d	650		0	Cognitive maps (Psychology)
d	650		0	Conceptual structures (Information theory)
b	710	2		ERIC Clearinghouse on Information & Technology.
s	830		0	ERIC digest (Syracuse, N.Y.) :|vEDO-IR-97-5.
y	856	4	1	|uhttp://library.du.edu/findit/peak/redirect.cfm?LinkURL=http://www.ericfacility.net/ericdigests/ed407938.html|z
				Access online version

ColdFusion database keeps track of click-throughs

UNIVERSITY OF DENVER
PENROSE LIBRARY

Penrose Home / Finding What You Need / Services / About the Library / What's New / Management Menu / Log Out

Peak Redirect Links
These are all the peak links that library users have used.

Archive These Records

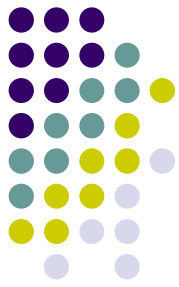| Date	IP	Website
01/21/04	13:30	216.150.202.33	http://www.socialsecurity.gov/employer_info/evs2000.html
01/21/04	13:34	130.253.5.179	http://purl.access.gpo.gov/GPO/LPS39895
01/21/04	13:42	130.253.5.138	http://purl.access.gpo.gov/GPO/LPS34979
01/21/04	15:53	65.102.206.97	http://nces.ed.gov/pubs/97402.html
01/21/04	16:00	130.253.5.179	http://purl.access.gpo.gov/GPO/LPS35848
01/21/04	16:08	130.253.5.179	http://purl.access.gpo.gov/GPO/LPS35848
01/21/04	16:09	130.253.5.179	http://purl.access.gpo.gov/GPO/LPS35848
01/21/04	16:21	130.253.32.162	http://purl.access.gpo.gov/GPO/LPS39266
01/21/04	16:27	130.253.32.162	http://purl.access.gpo.gov/GPO/LPS33516

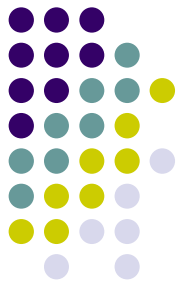# Access database enhances data

URL Management

- The new system has changed our updating process
- Reactive, rather than proactive
- URLs usually fixed within one day

Results After One Year

Results by Agency

B

B 002.001	1	International Broadcasting Bureau (1994-)

C

C 001.001	14	Department of Commerce (1913-)
C 001.202	2	Technology Administration
C 003.001	142	Bureau of the Census (1903-1972, 1975-)
C 013.001	7	National Institute of Standards and Technology (1934-)
C 021.001	3	Patent and Trademark Office (1975-)
C 055.001	13	National Oceanic and Atmospheric Administration (1970-)
C 055.101	1	National Weather Service (1970-)
C 055.201	8	National Environmental Satellite, Data, and Information Service (1970-)
C 055.301	2	National Marine Fisheries Service (1970-)
C 055.401	2	National Ocean Service (1970-)
C 059.001	10	Bureau of Economic Analysis (1975-)
C 060.001	8	National Telecommunications and Information Administration (1978-)
C 061.001	21	International Trade Administration (1980-)
C 063.001	1	Bureau of Industry and Security

CC

CC 001.001	2	Federal Communications Commission (1934-)

CR

CR 001.001	29	United States Commission on Civil Rights (1957-)

D

D 001.001	28	Department of Defense (1949-)
D 002.001	3	American Forces Information Service (1977-)
D 005.001	12	Joint Chiefs of Staff (1949-)
D 005.401	46	National Defense University (1977-)
D 012.001	1	National Guard Bureau (1958-)
D 015.001	1	Defense Threat Reduction Agency (1959-)
D 101.001	146	Department of the Army (1949-)

J

J 001.001	78	Department of Justice (1870-)
J 016.001	2	Federal Bureau of Prisons (1930-)
J 021.001	4	Immigration and Naturalization Service (1940-)
J 024.001	2	Drug Enforcement Administration (1973-)
J 026.001	16	Bureau of Justice Assistance
J 027.001	1	United States Parole Commission
J 028.001	97	National Institute of Justice (1979-)
J 029.001	23	Bureau of Justice Statistics (1979-)
J 032.001	81	Office of Juvenile Justice and Delinquency Prevention (1974-)
J 034.001	15	Office for Victims of Crime
J 035.001	6	Violence Against Women Office (1995-)
J 036.001	8	Office of Community Oriented Policing Services
J 037.001	2	Office of the Inspector General
L 036.001	1	Employment Standards Administration (1971-)

JU

Ju 006.001	3	Supreme Court (1790-)
Ju 010.001	30	Administrative Office of the United States Courts (1939-)
Ju 013.001	2	Federal Judicial Center (1967-)

L

L 001.001	10	Department of Labor (1913-)
L 002.001	12	Bureau of Labor Statistics (1913-)
L 029.001	3	Bureau of International Labor Affairs (1959-)
L 035.001	1	Occupational Safety and Health Administration (1971-)
L 036.101	2	Women's Bureau (1971-)
L 036.201	2	Wage and Hour Division (1971-)
L 037.001	5	Employment and Training Administration (1975-)
L 041.001	1	Office of Disability Employment Policy

LC

LC 001.001	11	Library of Congress (1800-)
LC 002.001	1	General Reference and Bibliography Division (1944-)
LC 003.001	2	Copyright Office (1897-)
LC 009.001	7	Descriptive Cataloging Division (1940-)
LC 014.001	3	Congressional Research Service (1914-)

Results by Agency, cont.

Y 003.Ad 9/1	1	United States Advisory Commission on Public Diplomacy
Y 003.Am 3/	4	American Battle Monuments Commission (1923-)
Y 003.B 27	1	Architectural and Transportation Barriers Compliance Board (1976-)
Y 003.C 73/5	2	Commodity Futures Trading Commission (1974-)
Y 003.C 76/3	1	Consumer Product Safety Commission (1972-)
Y 003.D 36/2	2	Defense Base Closure and Realignment Commission
Y 003.D 63/3	2	National Council on Disability
Y 003.El 2/3	1	Federal Election Commission (1974-)
Y 003.Eq 2	5	Equal Employment Opportunity Commission (1965-)
Y 003.F 31/0	2	Federal Deposit Insurance Corporation (1933-)
Y 003.H 62	1	Advisory Council on Historic Preservation (1966-)
Y 003.H 75	2	Interagency Council on the Homeless
Y 003.In 8/25	1	Inter-American Foundation (1969-)
Y 003.L 71	5	National Institute for Literacy (1991-)
Y 003.M 46/3	1	Medicare Payment Advisory Commission (1987-)
Y 003.N 21/2	1	Corporation for National and Community Service
Y 003.N 88	4	Nuclear Regulatory Commission (1975-)
Y 003.P 31	84	United States Institute of Peace (1984-)
Y 003.P 38/2	1	Pension Benefit Guaranty Corporation (1974-)
Y 003.Se 5	2	United States Sentencing Commission (1987-)
Y 003.Se 5/2	2	Commission on the Bicentennial of the U.S. Senate
Y 003.T 22/2	12	Office of Technology Assessment (1972-1995)
Y 003.zzz002	13	Commissions, Committees, and Boards

Y 4

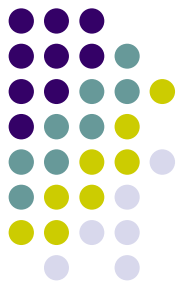|		
Y 003.F 76/2	1	Foreign Economic Policy Commission (1953-1954)
Y 004.Ag 4	7	Special Committee on Aging (Senate)
Y 004.Ag 8/1	1	Committee on Agriculture (House)
Y 004.Ag 8/3	2	Committee on Agriculture, Nutrition, and Forestry (Senate)
Y 004.Ap 6/2	25	Committee on Appropriations (Senate)
Y 004.Ar 5/2	1	Committee on Armed Services (House)
Y 004.As 8	1	Select Committee on Astronautics and Space Exploration (House)
Y 004.B 22/3	13	Committee on Banking, Housing and Urban Affairs (Senate)
Y 004.B 85/3	4	Committee on the Budget (House)

Y 004.C 44	21	Joint Special Committee on Chinese Immigration
Y 004.C 73/7	17	Committee on Commerce, Science, and Transportation (Senate)
Y 004.C 73/8	14	Committee on Energy and Commerce (House)
Y 004.Ed 8/1	15	Committee on Education and the Workforce (House)
Y 004.En 2	8	Energy and Natural Resources Committee (Senate)
Y 004.F 49/2	12	Committee on Financial Services (House)
Y 004.F 76/1	1	Committee on Foreign Affairs (House)
Y 004.F 76/2	44	Committee on Foreign Relations (Senate)
Y 004.G 74/7	64	Committee on Government Reform (House)
Y 004.G 74/9	25	Committee on Governmental Affairs (Senate)
Y 004.H 81/3	2	Committee on House Administration
Y 004.In 2/02	1	Committee on Indian Affairs (Senate)
Y 004.In 2/11	6	Select Committee on Indian Affairs (Senate)
Y 004.In 8/16	53	Committee on International Relations (House)
Y 004.In 8/18	1	Select Committee on Intelligence (House)
Y 004.J 89/1	15	Committee on the Judiciary (House)
Y 004.J 89/2	85	Committee on the Judiciary (Senate)
Y 004.L 11/4	27	Committee on Health, Education, Labor, and Pensions (Senate)
Y 004.P 93/1	1	Joint Committee on Printing
Y 004.P 96/1	10	Committee on the Environment and Public Works (Senate)
Y 004.R 31/3	15	Committee on Resources (House)
Y 004.Sci 2	5	Committee on Science (House)
Y 004.Se 2	11	Commission on Security and Cooperation in Europe
Y 004.Sm 1	3	Committee on Small Business (House)
Y 004.T 68/2	3	Transportation and Infrastructure Committee (House)
Y 004.V 64/4	1	Committee on Veterans' Affairs (Senate)
Y 004.W 36	15	Committee on Ways and Means (House)

Top Agencies

Congressional Hearings/Prints (Y 4)	523
Congressional Reports/Documents (Y 1)	409
Justice Dept. (J)	333
General Accounting Office (GA)	308
Defense Dept. (D)	288
Heath and Human Services (HE)	250
Commerce Dept. (includes Census) (C)	227
National Archives (AE)	195
Education Dept. (ED)	177
Congressional Commissions (Y 3)	147
Executive Office of the President (PREX)	116
Interior Dept. (I)	89
State Dept. (S)	68
Agriculture Dept. (A)	63
National Academy of Sciences (NA)	62
Environmental Protection Agency (EP)	46
Library of Congress (LC)	39

Top Titles

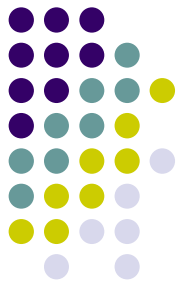| SuDocs	CountOfSu	Title
Y 1.1/8:107-555/ PT.1	112	Arming Pilots against Terrorism Act : report
Na 1.8:	75	Proceedings of the National Academy of Sciences of the United States
C 3.134:	37	Statistical abstract of the United States
PREX 3.19:	26	Studies in intelligence
HE 20.30:	24	Public health reports
PREX 2.6/2:IN 27/987	23	Standard industrial classification manual, 1987
AE 2.106:	17	Federal register
JU 10.8:	15	Federal probation
AE 2.114:	15	Public papers of the presidents of the United States
J 1.14/8:	15	FBI law enforcement bulletin
AE 2.109:	15	Weekly compilation of Presidential documents
Y 4.J 89/2:S.HRG.106-900	14	Juvenile intervention programs that work : hearing
HE 20.3559:	13	Environmental health perspectives : EHP
I 19.127:2004-3009	12	Potential effects of individual sewage disposal system density on ground
Y 4.J 89/2:S.HRG.106-1060	12	Music on the Internet : is there an upside to downloading? : hearing
AE 2.106/3:	11	Code of federal regulations
PREX 2.6/2:IN 27/2002/CLOTH	11	North American industry classification system : United States, 2002
AE 2.108/2:	11	The United States government manual

Click-throughs by Date

Date	Total Accesses	URL not in CGP
1824	2	2
1837	1	1
1854	2	2
1860	1	1
1861	5	5
1863	3	3
1866	2	2
1873	2	2
1875	6	6
1876	4	4
1877	1	1
1878	4	4
1880	1	1
1881	1	1
1889	1	1
1898	3	3
1901	3	3
1903	3	3
1908	3	3
1911	2	2
1918	1	1
1919	1	1
1922	4	4
1941	2	2
1944	1	1
1959	1	1
1961	1	1
1963	1	1
1965	3	3
1967	1	1
1970	6	6
1971	3	3

Date	Total Accesses	URL not in CGP
1973	2	1
1975	1	1
1976	1	1
1977	4	2
1978	1	1
1982	2	2
1983	9	1
1984	15	14
1985	3	3
1986	9	8
1987	30	29
1988	11	10
1989	24	16
1990	22	13
1991	23	11
1992	18	9
1993	32	21
1994	87	52
1995	112	76
1996	126	72
1997	187	106
1998	234	154
1999	210	101
2000	338	76
2001	366	30
2002	610	27
2003	965	85
2004	292	90
Total	**3809**	**1087**

These results tell us that users will access older documents if they are available online: Evidence that the GPO digitization of legacy docs will be well-worth the effort.

Circs vs. Online Access

SuDocs	Circulated	Online
A	54	63
AC	1	1
AE	6	195
B	0	1
C	47	227
CC	2	2
CR	4	28
D	85	288
E	6	37
ED	220	177
EP	35	46
FEM	0	3
FR	0	6
FS	14	0
FT	11	23
GA	8	308
GP	2	37
GS	1	24
HE	172	250
HH	18	17
HS	0	6
I	189	89
ITC	2	5
J	42	333
JU	0	35
L	32	35
LC	8	39
MS	0	1
N	0	7

SuDocs	Circulated	Online
NA	2	62
NAS	31	27
NF	0	9
NS	0	6
P	0	2
PE	0	5
PM	0	1
PR	39	21
PREX	47	116
PRVP	0	3
S	34	68
SBA	6	9
SE	0	4
SI	20	4
SSA	0	6
T	11	17
TD	27	30
VA	4	6
W	0	22
X	0	20
Y 1	364	409
Y 3	46	147
Y 4	602	523
Y 10	2	7
Z	0	2
Totals	**2080**	**3809**

Most online accesses predictably follow circs; others do not – because of circ restrictions (ex.: AE) or mf formats (ex: GA).

Benefits

- Informs book purchases
- More accurate item selection
- Retro cataloging can follow clickthroughs
- Selective URL maintenance: saves time
- Value of aggressing adding projects
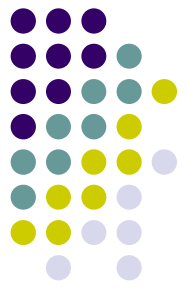
Aggressive adding Worth it?: YES

- *Fully 39% of all clicks involved URLs added not by the GPO, but by our documents staff.*

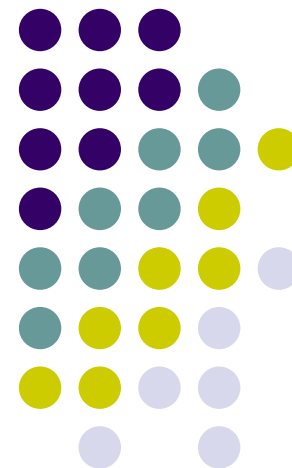# These Results Forthcoming:

- *Online Information Review*, November, 2004

Access to Government Documents: Tracking a Library's Website

Federal Depository Library Conference
Washington, DC
Susan Xue
University of Colorado at Boulder
susan.xue@colroado.edu

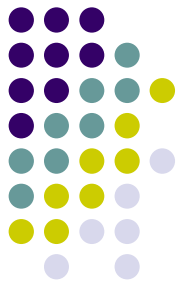# Tools of evaluating access to government documents

- Number of circulation
- Number of reference questions
- Number of times of viewing electronic titles in local online catalog
- Remote access to federal documents through depository libraries (GPOAccess)
- Number of visits to website

Why need a website?

- Currently 65% of GPO cataloged publications are online, 95% online in five years

- About 40% of these publications are links to agency websites (Judy Russell, 2003)

- There are many other publications on agency websites. A website could facilitate access to this huge volume of information

- Electronic government publications occupy more than 85% of the "deep web", search engines cannot reach

Government Publications Library, University of Colorado at Boulder

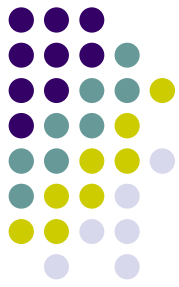http://ucblibraries.colorado.edu/govpubs/index.htm

What web statistics tell us (how many, where, and visited what)

- Number of visits
- Length of visits
- Top pages visited
- Visitors and demographics
- How do people find our site (search engines)
- What search terms lead people to our site
- What browsers are used
- Most downloaded files
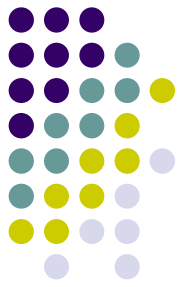
Profile of visits

07/01/03-06/30/04

- Hits (successful): 891,188
- Visits: 424,820
- Average visits per day: 1,160
- Average visit length: 00:11:22
- Visitors who visited once: 195,909
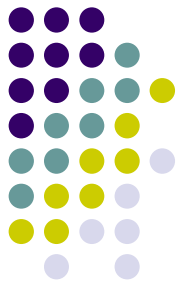- Visitors who visited more than once: 28,761

Top pages visited

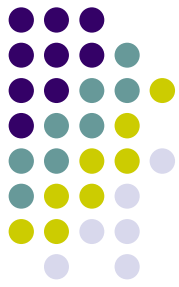| Pages	Visits	Avg. Time Viewed
1 Homepage	51,819	00:07:58
2 Federal government index page	48,019	00:01:19
3 Affirmative Action and Civil Rights	39,968	00:03:54
4 Guides to Public Records	26,061	00:04:30
5 Black history	25,687	00:04:24
6 Native American Treaties and Information	25,560	00:03:54
7 United States Maps and Atlases	17,371	00:02:33
8 Demographics and Statistics	15,408	00:03:27
9 Colorado State Resources	14,172	00:01:46
10 Aerial Photography and Satellite Imagery (foreign countries)	11,165	00:03:40

Most downloaded files

File	No. of Downloads
1 Europe from A-Z –guide to Europe	1,559
2 The ABC of Community Law	1,381
3 Questions of Jurisdiction	1,024
4 Government Publications Overview	510
5 Guide to use of Netscape Browser	507
6 A guide to environmental studies sources	476
7 A guide to League of Nation documents	189

Most active cities

City	Visits
1. Boulder, Colorado	43,713
2 Santa Clara, California	20,644
3 Chantilly, Virginia	14,908
4 Denver, Colorado	11,898
5 Atlanta, Georgia	7,993
6 Chicago, Illinois	7,183
7 San Mateo, California	6,668
8 New York, New York	6,415
9 Washington, D.C.	5,658
10 San Jose, California	5,527

Most active countries

Country	Visits
1.United States	370,774
2 Canada	10,065
3 United Kingdom	8,532
4 Australia	3,162
5 Germany	2,269
6 France	1,961
7 Netherlands	1,813
8 China	1,808
9 Japan	1,449
10 India	1,440

Top referring sites

Site	Visits	
1 No Referrer	132,733	**31%**
2 http://www.google.com	114,784	**27%**
3 http://search.yahoo.com	47,152	**11%**
4 http://search.msn.com	31,043	**7%**
5 http:// www-libraries.colorado.edu	16,222	**4%**
6 http://aolsearch.aol.com	6,008	
7 http://web.ask.com	5,531	
8 http://www.google.ca	4,857	
9 http://www.doi.gov	3,998	
10 http://search.colorado.edu	2,446	

Top search phrases leading people to our site

Phrases	Phrases found	% of Total
1 affirmative action	17,056	7.22%
2 united states government	4,227	1.79%
3 colorado public records	3,618	1.53%
4 alcohol statistics	2,374	1.00%
5 citizenship test	2,359	0.99%
6 black history month	1,794	0.75%
7 government acronyms	1,764	0.74%
8 public records	1,750	0.74%
9 united states map	1,525	0.64%
10 native american treaties	1,474	0.62%

…Top search phrases

11	satellite photos	1,240	0.52%
12	navy acronyms	1,018	0.43%
13	international organizations	1,006	0.42%
14	congressional research service	981	0.41%
15	united states goverment	898	0.38%
16	united states maps	861	0.36%
17	black history	807	0.34%
18	affirmative actions	750	0.31%
19	affirmative action	730	0.30%
20	"affirmative action"	716	0.30%

A useful information delivery tool

- Web visits VS traditional access

	No of circ.	No of reference	No of visits
2001/02	19,681	8,320	
2002/03	17,698	8,892	310,554
2003/04	13,107	6,422	424,820

- People seem find information useful as average visit time was 11 minutes
- The website attracts visitors nationwide, especially people referred by search engines. 15% of visits were from Colorado, 87% were from the U.S., and 13% international.

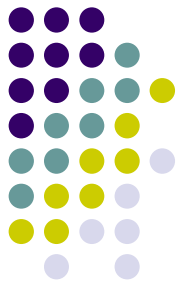# Top pages visited and top search phrases

- Top pages/search phrases are not always the ones we expected

- Possible explanations:

  - most web surfers only go through first two search result pages

  - too many websites on legislation and statistics…

  - too few websites on affirmative action, public records…

- Possible developments

  - make links to legislation/statistics pages more visible

  - develop the most accessed subject pages into more comprehensive resources

Develop more guides /instructions to resources

- Most downloaded files indicate that guides and instruction materials are popular
- Review and update existing documents
- Develop new guides/instructions
- Put a link to guides on the homepage

Role of search engines

- 64% of total visits were referred by search engines

- Google (61% of referral) search technology – page rank + keywords

- Other search engines – keywords + domain name + other

Impact of search engine technology on homepage design

- Homepage, as a likely recipient of high page rank, should be constructed accurately and informatively

- The most important links should be put on the first pages

- The architecture of the site should be shallow and consistent (Yahoo homepage is a model)

(Thelwall, 2002)

Government Publications Library Website (August 2002)

Resources

e-Resources: A-Z List

Ask an Expert

What's New!

State and Local Government	United States Government	Foreign Countries and International Organizations
Colorado	United States	Foreign Countries
Colorado by the Numbers an online statistical abstract of Colorado	CD-ROM Network (Government Publications Library access only.)	International Organizations, including:
Links to Information for All States	Technical Reports	• European Union • United Nations
Finding State, City, and County Websites	Attack on America	CD-ROM Network (Government Publications Library access only.)
Tax Forms	Tax Forms	

Government Publications Library Website (October 2003)

Most Used Databases

Campus access only
Seek staff assistance in Government Publications
Access UN
American FactFinder (Census)
American Memory
Bureau of Justice statistics
CIA world factbook
Colorado by the numbers
CQ electronic library
Declassified documents reference service
Defense technical reports
Digital national security archive
Energy citations
EUR-Lex
GPO Access
LexisNexis academic
LexisNexis congressional
....

Government Resources on the Web

United States

United States
CD-ROM network
Technical reports
Federal tax forms
War against terrorism
Ask an expert!

State Government

State of Colorado
Colorado by the numbers
Colorado tax forms
Links to information for all states
State, city, and county directories

Foreign Countries

Foreign countries
Foreign government links
Statistics

International Organizations

....

Other Resources

Search Tools

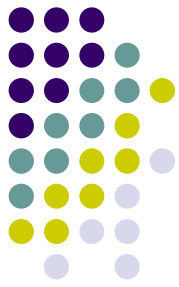CHINOOK ONLINE CATALOG

GoogleUncleSam
Other search engines

Hot Topics

Congressional reports on 9/11 intelligence
Investigation into the death of Dr. Kelly
North Korea
Columbia accident report

Research Guides

Collection overview
Guides to resources
Citing Internet resources
Model United Nations
....

Before and after the redesign (Oct. 2003)

Total visits		Average per day	Visits to homepage
2002/03	310,554	848	73,922
2003/04	424,820	1,160	110,793

A continuing task

- Monitor digitization projects by GPO
- Monitor digitization projects by agencies
- Monitor the development of technology
- Get to know publications on agency websites

Interested in reading more on website evaluation?

Online Information Review

Volume 28 Number 3, 2004

- Focusing on website evaluation, retrieval of online database, and search engines