Knowing Where They Went: Six Years of Online Access Statistics via the Online Catalog for Federal Government Information

Christopher C. Brown

As federal government information is increasingly migrating to online formats, libraries are providing links to this content via URLs or persistent URLs (PURLs) in their online public access catalogs (OPACs). Clickthrough statistics that accumulated as users visited links to online content in the University of Denver's library OPAC were gathered over a six-year period and were analyzed. Among the conclusions were that DU users prefer online content over print for both newer and older documents and that there is great benefit in adding URLs above and beyond the URLs supplied by GPO cataloging.

This article is a follow-up to "Knowing Where They're Going: Statistics for Online Government Document Access through the OPAC" published in *Online Information Review*.[1] That article discussed the genesis of the University of Denver's (historically abbreviated DU) initiative to track online clickthroughs of U.S. federal documents accessed via the online public access catalog (OPAC). Now, with six years of data in the rearview mirror, it is possible to analyze the data and make some meaningful statements regarding trends, cataloging practices, and depository administration. Clickthrough data are important for several reasons: 1) they provide a consistent, regularized view of usage, accessed via a singular technology (the OPAC); 2) they are the closest com- parative statistic to circulation statistics for tangible items; and 3) they provide a view of usage over time. As with circula- tion statistics, these data do not indicate whether users actually used or read the materials. However, these statistics at least provide us a relative guide to usage of online items.

The digital age for U.S. federal docu- ments began in 1994 when the Govern- ment Printing Office (GPO) introduced the Web portal GPO Access, providing the public systematic access to selected government documents online.[2] By 2002, nearly 60 percent of new Federal De- pository Library Program (FDLP) titles were available online.[3] That percent had increased to about 97 percent by 2009.[4] In large depository libraries, the space allocation for print documents collec-

Christopher C. Brown is Associate Professor and Reference Technology Integration Librarian at the University of Denver, Penrose Library; e-mail: cbrown@du.edu. © Christopher C. Brown

tions is substantial. Document shelving areas are gradually being replaced with group study, open spaces, and other programmatic usages. It was noted in a recent survey that Association of Research Libraries' directors tend to prefer digital access.[5] With these factors in mind, library directors of the Colorado Alliance of Research Libraries (the Alliance) requested that Colorado's depository libraries try to reduce the redundancy of holdings. This goal could be achieved in various ways including deselection of GPO items with tangible formats, weeding of collections, and placing materials in storage.

To understand the extent of redundancy in Colorado, depository libraries within the state measured the linear feet of their print collections by Superintendent of Documents (SuDocs) call number class in early 2008. Since selective depositories receive more or less the same titles based upon their item selection profiles, a measurement of linear feet proves to be a true reflection of extent of redundancy across area depositories. The table below shows the extents of selected collections.

The Alliance libraries are approaching the redundancy issue in varying ways. The University of Denver has chosen to keep nearly all of its "legacy" collection of older documents, while at the same time reducing selections of items with tangible receipts (print, fiche, or computer disks) to virtually nothing. Emphasis is

placed on online formats by subscribing to Marcive's Documents without Shelves subscription service.[6] With this service, libraries can receive electronic-only records created by GPO, as well as the option to subscribe to the changed records service (that is, changes in publication history, URLs, or corrected records). Monthly record distribution can be based on a specific item selection, or (as in the case of the University of Denver) libraries can elect to receive 100 percent of electronic items. In addition, the library is in the midst of a multiyear project to add URLs to older documents.

This article surveys the URL click-through tracking initiative undertaken by the University of Denver over the past six years, compares the results of the online usage statistics with usage of print, and discusses the implications to depository libraries.

Acquiring, adding, and maintaining URLs in document catalog records raises a series of questions. Are users accepting electronic access as the format of preference, or are they still requesting documents in tangible formats? Is it worth the efforts as measured by ever-increasing online access statistics to add URLs to older documents? Is it worth the money and time to purchase records for large collections such as digital hearings or the U.S. Congressional Serial Set? What URL maintenance strategies make the

TABLE 1		
Documents Linear Feet in Selected Colorado Depositories		
Institution	**Depository Type**	**Linear Feet of Print Documents**
University of Colorado at Boulder	Regional (100%)	20,051
Denver Public Library, Denver	Regional until June 30, 2009, Now a large selective	24,000
Colorado College, Colorado Springs	Selective (33%)	6,464
Colorado School of Mines, Golden	Selective (43%)	7,440
Colorado State University, Fort Collins	Selective (69%)	20,223
University of Denver, Denver	Selective (70% until recently; now 5%)	10,395

best use of time and personnel resources? Another question is interesting, but is beyond the scope of this study: Are online links adequate substitutes for print publications?

Background of the University of Denver, Penrose Library

DU is a private not-for-profit university with just under 10,000 students, roughly half graduate and half undergraduate students. Penrose Library is the main campus library, with an administratively separate law library across the street. References to "the library" in this article are meant to include only Penrose Library. The university became a federal depository library in 1909. In recent years the library maintained a large selection rate of nearly 75 percent. With the construction of a shared Harvard-style remote storage facility in 2000, Penrose Library began placing materials, including about 80 percent of its sizeable documents collection (10,395 linear feet of paper documents), in remote storage. To compensate for much of the documents collection being offsite, major emphasis was placed on ensuring online access to as much content as possible. This was accomplished in several ways: acquisition from Marcive of MARC records with GPO PURLs in the records, aggressive addition of URLs to older content as it was discovered, record loads of MARC records for large sets, such as the U.S. Congressional Serial Set from third-party vendors, and project-based URL additions.

In July 2003, the library started tracking online clickthroughs to all government documents accessed via the OPAC. This was accomplished by appending to every URL a prefix that redirects to a library server that collects date and time and URL accessed, then redirecting the user to the destination URL. Each evening the statistics are tabulated and any URL errors are corrected. After six years of this, the library has a rich data source of online accesses, URL error rates, and cross-tabulated totals.

Review of the Literature and Various Implementations

Much has been written about the transition to electronic formats and user acceptance of the electronic over tangible formats in the context of serials.[7] Users seem not to have taken to e-books as quickly as they have e-journals. A 2005 survey at the University of Denver showed that, although e-books were used by about half of the campus community, most use them only occasionally, with a slight preference for print over electronic format.[8] Online serials have been much more widely accepted by libraries and users than have e-books, as recently noted at Central Michigan University.[9] The corner was turned recently at the University of Colorado at Boulder, when, in Fiscal Year 2007–08, 56 percent of their materials budget was spent on electronic resources, as opposed to print.[10]

With widespread acceptance of online serials, and a less enthusiastic acceptance of e-books, what about documents usage? Mary Schneider Laskowski used a survey methodology to assess attitudes of document specialists and users of documents.[11] Her survey results from 1999 showed an increasing popularity of online access, while at the same time dissatisfaction with user interfaces was expressed. More recently, Jennie Burroughs found through surveys that users generally prefer to find and access government information online.[12] A report released in late 2009 by Ithaka S+R observed that users increasingly expect to discover information on their own, meaning that seamless discovery of online government information is imperative.[13]

While many have undertaken the survey method to assess user patterns and acceptance of online documents, several have pursued systematic tracking of "clickthroughs" to online documents. The first such endeavor was the 2003 implementation the author developed at the University of Denver,[14] which this study is summarizing.

At least six other libraries in addition to the University of Denver have been track-

ing online clickthroughs to documents via the OPAC. Dr. Martin Luther King, Jr. Library is a combined library of San Jose State University and San Jose Public Libraries. Susan Kendall, after seeing the success with the University of Denver's link tracking, was able to have link tracking implemented in the San Jose State online catalog by assigning a unique ID number to each resource and appending that ID number into a URL prefix. Susan Kendall, Celia Bakke, and Lisa McDaniels presented a summary and analysis of their URL clickthrough results.[15] They noted that online clickthroughs closely follow the academic calendar and course research. Subsequent unpublished statistics from San Jose State University have shown increasing growth in document clickthroughs and an apparent online acceptance.[16]

Other libraries have also implemented URL tracking prefixes to document URLs, but no publications of their projects have been produced to date. As part of the consortial services it offers, the Louisiana Library Network (LOUIS) can, at request of specific member libraries, append a prefix before each document URL found in the member libraries' OPACs. Louisiana State University in Baton Rouge does this for federal documents, as apparently do other libraries in the consortium as well.[17] Auburn University Libraries has been tracking online NASA technical reports, ERIC documents, and the Department of Energy's Office of Scientific and Technical Information (OSTI) documents with its prefix, and is now in the process of expanding this to all electronic documents.[18] The University of Colorado at Boulder is currently tracking selected congressional hearings with a prefix appended before its URLs. A publication comparing online access with print circulation is forthcoming.[19] Brigham Young University – Idaho (Rexburg) also appends a prefix before document URLs and uses these numbers to get an overall feel for how many documents are being accessed.[20] At the 2009 Fall Depository Library Conference, a panel presentation featured the clickthrough

projects of the University of Denver, San Jose State University, Auburn University, Louisiana State University, Brigham Young University – Idaho, and the University of Colorado at Boulder.[21] In mid-2009, Mercer University in Macon, Georgia, began tracking online clickthroughs to government documents in their online catalog.[22]

Since October 2000, the GPO has been providing a URL tracking system of sorts.[23] This service allows depository libraries to register Internet subdomains from which online links to PURLs within GPO Access resources originate (such as Web sites, research guides, and library online catalogs). There is a monthly tally of clickthroughs from each of these registered domains to GPO content. Although these results are more broadly based than just OPACs, they do not have the degree of granularity that the above-mentioned clickthrough projects are capable of: title, SuDocs number, and the like. Given these shortcomings, these statistics are of little value to depositories, other than comparing online accesses between depository libraries. Depository libraries generally find these data too broad and unable to provide the desired level of detail.[24]

Links to online documents in the OPAC come not only from GPO PURLs, but also from vendor-supplied records with individual links to online documents. The DU project tracked these clickthroughs as well as the freely available government documents links. Examples of these licensed resources include LexisNexis digital hearings, Readex Digital Serial Set, and ERIC records supplied by Marcive (although most of the ERIC documents themselves are not properly considered government documents, the index itself is a government endeavor of the Department of Education).

While the focus of this article is on tracking clickthroughs to government documents via the OPAC, much of the emphasis in the literature on URLs in the OPAC seems to be upon inaccuracy of the links ("linkrot"). For example, Gerald Burke, Carol Ann Germain, and Mary K.

Van Ullen surveyed 24 ARL libraries and found that one library had an error rate of up to 58.33 percent for URLs for freely available resources, including government documents.[25] They found that few libraries in their study had time to devote to systematic and proactive URL maintenance.

The 2004 *Online Information Review* article discussing the genesis of the project under discussion in this paper emphasized the expansive possibilities of large-scale additions of URLs to the OPAC.[26] DU had thousands of items in tangible formats in its OPAC, many of which also happened to be online, but there was no connection between these two worlds—the library world of print and microforms and the online world of full content. The extent to which the OPAC fails to tell the user of these linkages is the extent to which the catalog, the library, and indeed the library profession are perceptively and actually obsolete. This is the motivation for the author's aggressive addition of URLs to online content above and beyond the PURL links provided in GPO cataloging, as documented later. The focus of the 2004 article was not upon URL maintenance, but upon establishing a link-tracking system by placing a prefix to each URL that first of all redirects to a local library server; then records date, time, and URL; and finally (but almost instantly) redirects the user to the destination Web address. Statistical summary of one year of clickthroughs was provided in that study.

Nicholas Joint surveys libraries in the United States and the United Kingdom and discusses the addition of links to free Internet resources.[27] He observes that Scottish libraries are so concerned with invalid links that they tend to avoid adding links to free resources in their catalog records. He implies that the University of Denver takes "the radical path" of including freely available Web sites "that would normally be the preserve of Google and other search engines."[28] Joint generalizes about circulation of physical materials vs. online use via the OPAC and infers that

OPAC usage must be "in parallel" to (assumedly this means about the same as) circulation figures for tangible materials[29]. But since he admits that no figures are available for OPAC usage, his conclusions have little weight.

C. Rockelle Strader and Farrell D. Hamill focus more on why URLs are bad or become bad.[30] They present detailed analysis of error types, URL-checking protocols, and repair procedures—an approach that is typical of the very labor-intensive proactive URL management practice.

In addition to the 2004 article about the University of Denver documents clickthrough-tracking project, a subsequent article was published in 2008 discussing the expansion of the project to other freely available online content in the OPAC above and beyond federal documents.[31] The same URL tracking technology was extended to free online resources beyond U.S. federal documents, including publications in the following collections: the National Academies Press, Brookings Institution, Wright American Fiction, Human Rights Watch, Making of America, the RAND Corporation, University of California Press, and many others. Three years of cross-tabulated data show increasing access to online versions.

An article by Elizabeth Meagher and Christopher C. Brown in *Library Hi Tech* chronicles the decision-making processes at the University of Denver concerning adding URLs to bibliographic records, whether to use a single-record approach for online formats vs. multiple records (one for each distinct format), and URL maintenance issues.[32] The article focuses more on workflow issues and URL maintenance than it does on clickthrough data for free resources.

A Decade of URL Growth

At the University of Denver, it was not until 1998 that URLs began to appear in the online catalog, the result of initiating a subscription to GPO cataloging records via Marcive. As can be seen in figure 1, the number of URLs in the DU OPAC has

grown to over one million overall, with government documents laying claim to 45 percent of those links.

To understand how the University of Denver's government document links grew to such a large extent, it is necessary to explain the various projects the library has undertaken to aggressively add URLs to records beyond those supplied by the GPO.

Special URL-Addition Projects

The library wanted to provide online access to more resources than GPO provided through its MARC records. Because the GPO Access portal, the GPO PURL server, and GPO cataloging were not initially well integrated, PURLs were not incorporated into catalog records for many 1995 and 1996 materials. Even now a substantial amount of GPO Access content is not represented in GPO's Catalog of Government Publications and thus is not distributed through the various catalog record distribution services such as Marcive.

Beginning in 2000–2001, the DU library decided to provide access to this content using several automated processes to add URLs to Senate Documents (Y 1.1/3:), Senate Treaty Documents (Y 1.1/4:), Sen-

ate Reports (Y 1.1/5:), Senate Executive Reports (Y 1.1/6:), House Documents (Y 1.1/7:), House Reports (Y 1.1/8:), Public Laws (AE 2.110:), and GAO Reports (GA 1.13) from 1995 and 1996 onward.

The growth in URLs in the DU OPAC is attributable to a combination of the monthly loads of GPO records from Marcive, as well as the special project record loads that have been done over the years. Table 2 shows growth of documents URLs by fiscal year (July 1–June 30).

The following year, the library added links to its records for 7.5-minute topographic quadrangle maps. Maps are challenging to use online, so this is a clear case where the print format is easier to use. Nevertheless, as an experiment to see if linking a derived URL to Terraserver[33] was possible, and whether users would make use of the online versions, the library developed a process to extract the map boundaries from the MARC records, convert these coordinates from degrees, minutes, and seconds into decimal coordinates, and thereby derive the center point of each topographic map. This allowed derivation of the appropriate URL to dynamically generate the Terraserver

FIGURE 1
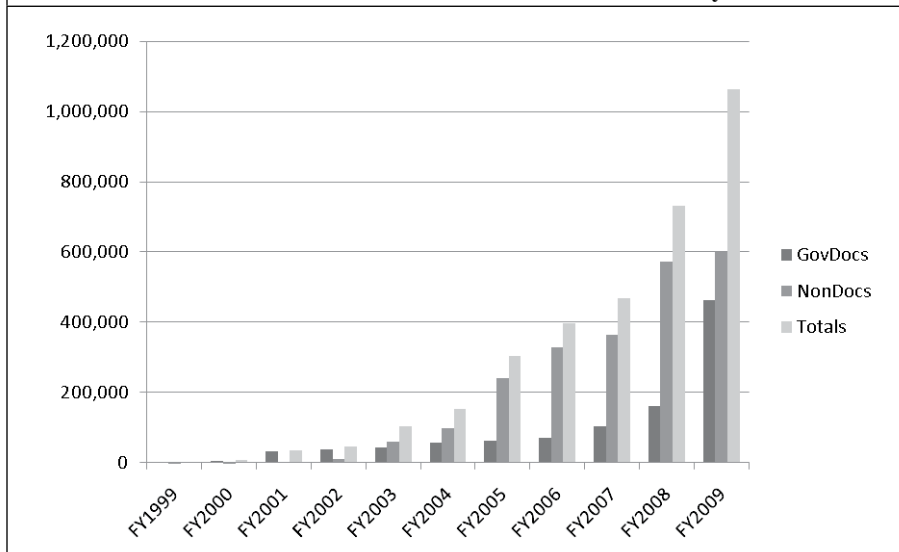**URL Growth in Government Documents at the University of Denver**

		TABLE 2
		Chronology of URL-Adding Projects at the University of Denver Penrose Library
Fiscal Year	**Docs URL Totals**	**Special Projects**
FY1999	1,848 docs URLs	• Marcive subscription began; URLs in docs records
FY2000	5,549 docs URLs	
FY2001	32,800 docs URLs	• Added thousands of records from GPO Access that GPO never added to their records • Added over 4,000 URLs to LOC Thomas legislative histories • Added 1,425 URLs to Economic Census records
FY2002	36,326 docs URLs	• Added 456 links to online 7.5-minute topographic maps
FY2003	43,307 docs URLs	
FY2004	55,508 docs URLs	• Began tracking URL clickthroughs • Added 207 URLs to U. of Md. Thurgood Marshall Law Library Civil Rights documents • Added 115 URLs to Law Library Microform Consortium Online
FY2005	62,374 docs URLs	
FY2006	69,341 docs URLs	
FY2007	103,021 docs URLs	• Added over 24,000 links to NASA technical reports
FY2008	159,543 docs URLs	• Added 40,690 URLs to LexisNexis digital hearings and committee prints • Added 9,535 URLs to older GAO records
FY2009	453,243 docs URLs	• Added additional 20,000 URLs to records for LexisNexis digital hearings and committee prints • Added 248,134 MARC records for Readex digital Serial Set • Added 12,902 OSTI MARC records • Added 5,913 URLs for public laws via HeinOnline

map and use an automated process to write the URLs into the MARC records.[34] This project has received accolades from our many remote users.

In FY2007 major local URL additions were made to the University of Denver's NASA technical reports records. These records were for microfiche versions of technical reports that had recently become accessible through the NASA Technical Reports Server. Although the GPO is slowly adding or updating records for older NASA reports, it seemed that this could happen much faster locally with some automation. The author downloaded all of DU's NASA microfiche records to a desktop computer, then downloaded various bibliographic elements into an Access database: bibliographic control number, OCLC number, SuDocs number, and title (field 245 subfield a). From the SuDocs number, a URL to each NASA Technical Reports Server record could be derived. The author then checked these URLs to ensure that the link to full text actually existed. Numerous adjustments needed to be made for serials, multivolume works, and other anomalies. Then the appropriate URLs were written to the MARC records and the records were reloaded into the local catalog. In all,

over 24,000 NASA URLs were added. The process of verifying the URLs took four months, while the actual writing of the URLs to the records and reloading records into the ILS took several hours.

In FY2008, the author applied the same process to older General Accounting Office (GAO)[35] reports, with over 9,500 URLs added. In the same fiscal year, the library purchased the LexisNexis complete digital congressional hearings collections as well as the Congressional Research Digital Collection (CRDC) collection, which includes congressional committee prints. Since the library's hearings were already in remote storage, it was important to provide users with instantaneous access to the hearings, so waiting for the LexisNexis records was not an option. To date, about 60,000 URLs have been added to existing print records for digital hearings and committee prints.

FY2009 had a number of URL addition projects. Our catalogers loaded the first half of the Readex Serial Set records (over 248,000 records in all, covering 1817 to 1948) to the catalog in October 2008. This greatly increased document exposure in the catalog, especially considering that this rich research set was little used in paper form over many decades.

Also in FY2009, the library loaded a subset of the recently released records from the U.S. Office of Scientific & Technical Information (OSTI). These records can be dynamically generated and downloaded from the OSTI database (www.osti.gov/marcrecords); but, because of field-mapping issues from the OSTI index to MARC format, they present numerous problems for catalogers. Nevertheless, the library was able to load over 19,000 of the more than 200,000 existing records into the local catalog and track clickthroughs.

Finally, using automation processes similar to those used in the NASA and GAO projects, the author added URLs to 5,919 records for public laws linking out to full-text content in HeinOnline.

With this project background in mind, an explanation of the project methodology and presentation of results is in order.

Methodology for This Project

Rather than conducting another survey, useful as that might be, this study examines actual user clickthroughs to online federal government document content. Although users may discover this content through any number of means, including general Web search engines, this study tracks clicks that have been accessed via the library OPAC. A clickthrough script was appended to any and all document links (field 856) in the OPAC. This provides a regularized, rich source of data that can be compared with itself and with print circulation statistics. It should be noted that duplicate clicks by the same user and administrative clicks (such as testing done by cataloging and documents staff) have been eliminated from the data set.

The tracking procedures were chronicled previously in detail,[36] but a summary of the implementation is helpful for the present readers. Using our Innovative Interface's Millennium global update function, we placed a URL tracking prefix in front of every government document URL in the 856 field to keep track of time/date of access and URL requested. This technology has made it possible to track detailed statistics of online documents accessed over the past six years. The URLs are then tracked and statistics are maintained in a Microsoft Access database. Each document clickthrough is checked daily and corrections are made, if necessary.

Summarization and cross-tabulation of data was maintained throughout the project within the Access database. Totals were possible by URL domain, Superintendent of Documents call number, title, time and date, and URL errors. Comparisons to physical circulation statistics were then all the more meaningful.

Six Years of Clickthroughs: Presentation of the Data

What can six years of clickthrough statistics tell us? After collecting the data and looking up the datapoints via the Access database, the following are data elements that can be tracked: title of document

requested, Superintendent of Documents (SuDocs) number, date and time of access, whether the user was off-campus or on-campus, date of publication, and URL of the document. These data elements can then be researched and cross-tabulated to yield more interesting results such as:

- Numbers of clickthroughs by day, week, month, year, or fiscal year;
- Numbers of clickthroughs by government agency;
- Numbers of clickthroughs by URL-addition project;
- Difference between clickthrough date and document date (this tells the age of the document relative to the date the document was accessed via clickthrough); and
- Error rate by fiscal year.

As seen in table 3, by FY2009 the number of clickthroughs had more than doubled from the first year of the project. The number of bibliographic records (Bib Recs) with URLs grew by a magnitude of 10, due in large part to the infusion of Readex Congressional Serial Set records in FY2009. Of the 7,660 online clickthroughs in FY09, the gradual addition of 248,134 Serials Set records in that year resulted in 249 clickthroughs to URLs in Serial Set records that did not exist in previous years. Given the difficulty and rarity of anyone accessing the print Serial Set for many years, this kind of accelerated access is stunning to the author. It seems that users are becoming increasingly accustomed to online access, as well as having much more content to access.

Cross-tabulation of SuDocs stems makes it possible to see annual accesses by major stem. Detailed data down to the title level are available, but it would take up too much space to report in this article. The following table shows the breakdown by SuDocs classes.

As we look at clickthroughs by SuDocs stem, the congressional hearings (Y 4) consis-tently receive the highest number of clickthroughs (over 20 percent throughout the six-year period). This is in general agreement with circulation trends for the print counterparts.[37] In the case of the University of Denver, the access pattern generally follows the circulation numbers in each SuDocs class. There are exceptions to this, however. The AE (National Archives) class is much higher in the online realm, since most of the print versions had been designated as noncirculating public laws and statutes. The spike in E (Energy) Department clickthroughs is directly attributable to the OSTI record load. GA (Government Accountability Office and NAS (NASA) increased usage is attributable to the respective retrospective URL projects described above and the likely preference of users to avoid fiche if online content is available. Congressional Reports and Documents (Y 1.1/3: through Y 1.1/8:), and their subsequent publication in the Serial Set (Y 1.1/2:), have generally not circulated, but the numbers for online usage are extremely high.

Assessing the URL-Addition Projects

After investing so much time and effort in the various projects described here, it is crucial to assess their success. Was it worth the time and effort to add thousands of URLs as described above? In other words, was there an appropriate return on investment? Using the URL tracking system makes it possible to assess the results.

TABLE 3			
Number of Clickthroughs by Fiscal Year			
Fiscal Year	Total Docs Bib Recs	Bib Recs with URLs	Clickthroughs to Docs
FY2004	358,215	43,307	3,809
FY2005	373,200	55,508	4,504
FY2006	388,610	62,374	4,686
FY2007	401,454	103,021	5,217
FY2008	429,122	159,543	6,342
FY2009	711,315	463,121	7,660

TABLE 4						
URL Clickthroughs by Fiscal Year						
(See appendix for meanings of SuDocs classes)						
Class	FY04	FY05	FY06	FY07	FY08	FY09	%/Total
A	63	81	132	89	118	117	1.86
AC	1	3		5	1	2	0.04
AE	195	178	158	300	533	514	5.83
B	1		6		2	8	0.05
C	227	166	176	274	271	258	4.26
CAB			1	2			0.01
CC	2	4	7	7	2	6	0.09
CR	28	44	28	12	31	17	0.50
D	288	269	376	289	277	347	5.73
E	37	83	95	93	129	671	3.44
ED	177	180	140	149	158	161	3.00
EP	46	49	100	156	177	112	1.99
FCA			3				0.01
FEM	3	5	7	6	11	4	0.11
FHF		1				1	0.01
FM			1			1	0.01
FP		1					0.00
FR	6	12	5	2	3	4	0.10
FS			1		2	1	0.01
FT	23	19	16	33	24	39	0.48
FTZ						2	0.01
FW						1	0.00
GA	308	465	445	498	633	652	9.31
GP	37	68	156	187	189	208	2.62
GS	24	15	27	22	84	20	0.60
HE	250	291	282	280	214	315	5.07
HH	17	24	10	6	8	18	0.26
HS	6	19	44	26	23	24	0.44
I	89	163	151	215	247	360	3.80
IA				4			0.01
IC		3			1	9	0.04
ID		2			3	12	0.05
ITC	5	8	12	15	8	10	0.18
J	333	261	204	145	226	246	4.39
JU	35	18	27	15	20	15	0.40
L	35	47	61	55	46	77	1.00

Class	FY04	FY05	FY06	FY07	FY08	FY09	%/Total
TABLE 4 **URL Clickthroughs by Fiscal Year** (See appendix for meanings of SuDocs classes)							
LC	39	43	83	43	44	62	0.97
LR		1	1				0.01
MS	1	1	1	5		2	0.03
N	7	7	7	2	2	4	0.09
NA	62	75	68	31			0.73
NAS	27	41	48	109	178	182	1.82
NC				1			0.00
NCU					2	1	0.01
NF	9	4	10	10	7	9	0.15
NMB		1			2		0.01
NS	6	11	6	3	12	2	0.12
OP		1	1	2	3	1	0.02
P	2	3	1	3	5	1	0.05
PE	5	12	5	6	3	5	0.11
PM	1	4	3	1	1	1	0.03
PR	21	36	30	45	35	49	0.67
PREX	116	131	123	134	98	163	2.37
PRVP	3	1	7		2	4	0.05
S	68	113	84	104	106	55	1.65
SBA	9	5	16	8	5	20	0.20
SE	4	6	2	1	6	10	0.09
SI	4	3	2	10	7	14	0.12
SSA	6	12	20	20	19	31	0.34
T	16	20	41	31	33	40	0.56
TD	30	38	24	21	42	55	0.65
TDA						1	0.00
VA	6	13	18	7	6	6	0.17
W	22	14	17	4	5	10	0.22
X	20	28	33	55	26	28	0.59
Y 1	408	394	243	316	485	674	7.82
Y 3	148	184	160	193	155	262	3.42
Y 4	524	807	950	1110	1558	1674	20.56
Y 6		1	1				0.01
Y 7						1	0.00
Y 10	7	15	8	53	53	58	0.60
Y 11			1	1		1	0.01
Z	2		1	3	1	2	0.03

The Projects are each described in table 5. The *URL Count* is the number of URLs existing in the University of Denver OPAC throughout the tracking time span. *Coverage Dates* refers to the document publication dates. *Tracking Time Span* refers to the number of months the URLs were in the OPAC with tracking prefixes. *URL Clicks* is the total number of clickthroughs, including duplicates. *Unique URL Clicks* are URL clicks with duplicate clickthroughs removed. And *% Unique Accessed* shows what percentage of distinct titles within the collection was accessed.

Topographic maps were frequently used relative to the total number in the collection, with over 16 percent of the maps accessed at least once. The NASA and GAO projects, each offering online alternatives to microfiche access, received many hits, although the ratio of unique hits to the number of titles available was small. The LexisNexis digital projects, necessary because of placement of hearings in remote storage, were popular with users. The massive record load of Readex Serial Set records provided analytical access for the first time to the Serial Set. Researchers were able to discover via the OPAC older materials that very few people previously would have had the time or patience to discover.

Do Older Documents Get Accessed?

The Access database kept track of the date of the clickthrough as well as the date of the documents (publication date). By subtracting the publication date from the clickthrough date, we can see the relative "date distance."

As older content is added to the catalog, users are able to find it and use it. Note that the percent of documents that were accessed and were over 10 years old has dramatically increased in the past six years. This is attributable to the fact that they are now discoverable via the online catalog and that users doing keyword searches stumble upon them. Most of the growth in annual clickthroughs comes from access to an increasing number of older documents over 10 years old. Based on these data, it appears that it is well worth the effort to add links to older online documents.

Circulation vs. Online Access

Circulation statistics for government documents during the period of this study have shown a slight decrease and can now be characterized as flat. Online click-

TABLE 5 Special URL Addition Projects						
Project	URL Count	Coverage Dates	Tracking Time Span	URL Clicks	Unique URL Clicks	% Unique Accessed
Topographic Maps	456	1991 – 2001	Sept. 2003 – June 2009	101	76	16.6%
NASA Technical Reports	24,825	1976 – 2001	April 2007 – June 2009	310	263	1.06%
GAO Reports (older)	9,559	1976 – 1999	Aug. 2007 – June 2009	184	161	1.68%
LexisNexis Digital Hearings/ Committee Prints	57,200	1850 – 1995	July 2007 – June 2009	1027	851	1.49%
Readex Digital Serial Set	248,134	1817 – 1948	Sept. 2008 – June 2009	239	205	0.08%
OSTI Reports	19,901	2002 – 2006	July 2008 – June 2009	476	375	1.88%

	FY04	**FY05**	**FY06**	**FY07**	**FY08**	**FY09**
TABLE 6						
Date Distance (Age of Document Relative to Date of Clickthrough)						
by Fiscal Year						
Total Clicks	3,809	4,504	4,686	5,217	6,342	7,660
Docs up to 10 years old	3,542	4,155	4,170	4,369	4,996	5,600
percent	*93.0%*	*92.3%*	*89.0%*	*83.7%*	*78.8%*	*73.1%*
Docs over 10 years old	267	349	516	848	1,346	2,060
percent	*7.0%*	*7.7%*	*11.0%*	*16.3%*	*21.2%*	*26.9%*

throughs, however, have been trending upward. In 2004, online usage was 1.65 times greater than physical circulation. In 2008, online usage was 4.75 times greater.

Throughout the time of the project, print items circulated an average of 1,552 times per year. But the online access continues to trend in the upward direction, from 3,809 clickthroughs in FY2004 to 7,660 clickthroughs in FY2009.

It is important to state that print document collections should not be eliminated entirely. However, given user preferences for online content, and the need to free up library collections space for other uses, print holdings within a region can be consolidated and redundancy can be reduced as we note user trends of preferring online content.

The Free vs. the Fee-based

The emphasis of the FDLP is access to freely-available government information. However, depositories, and especially regionals, are highly encouraged to augment their digital federal collections with commercial online content.[38] This is the approach the University of Denver has taken, by providing links to commercial document content from LexisNexis, Readex, and HeinOnline in its MARC records. The clickthrough statistics can tell us if users are, in fact, making use of these products via the OPAC. With most of DU's documents in remote storage, we rely on clickthrough statistics to demonstrate that we are providing immediate access storage items.

FIGURE 2
Comparison of Online Access with Physical Circulation of Documents
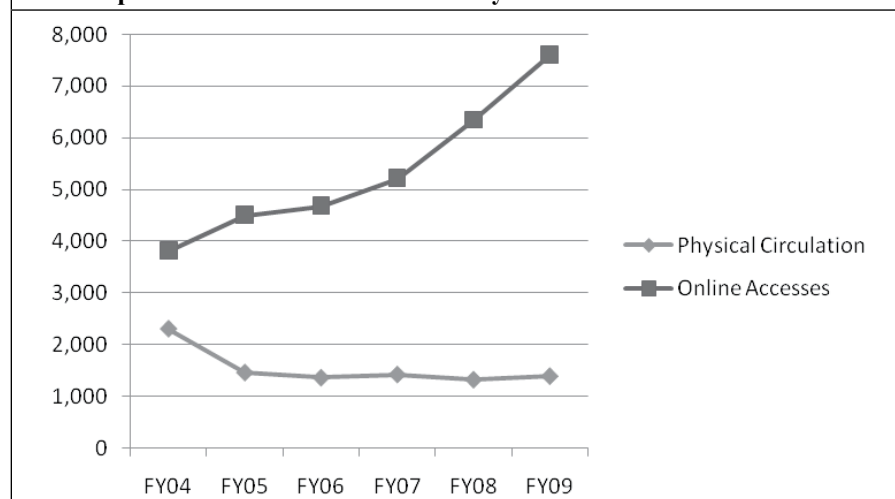
TABLE 7		
URL Clickthroughs for Selected Special Projects		
Product	**FY2008**	**FY2009**
LexisNexis Digital Hearings/ Committee Prints	360	665
Readex Digital Serial Set	n/a	239
HeinOnline (various titles)	150	114

Table 7 shows use of direct links to content discovered via the OPAC, as opposed to requiring users to initiate research from each of the publisher portal sites. Most popular of the special projects has been the digital hearings. Even though LexisNexis now provides MARC records for individual titles, it was important to us in 2007 to add links to existing catalog records to save the time of the user and to get appropriate return on the university's investment. The clickthrough statistics bear out that adding these 60,000 links was well worth the time. The digital Serial Set, for the first time searchable in detail via the local online catalog, has also been a success. In addition to standard HeinOnline content such as the Code of Federal Regulations and the Federal Register, the decision to add individual records for public laws in HeinOnline has also proven successful. In the future, the library plans to add links to "slip treaties" as well.

Error-checking Models: Proactive vs. Reactive

The literature shows that error checking is of great concern to many librarians. This makes sense; our profession has always taken pride in creating and maintaining accurate records of publications. Many ILS' have built-in error-checking programs. There are inherent problems with using them, however. In the case of most government documents currently supplied through the GPO, PURLs are used. PURLs are, by definition, redirects. It would be a waste of time and resources to even check these, since the checking software will report these links as redirects, usually without checking to see if the redirect works. The newer Handle System, to which GPO will soon be migrating, also uses redirection technology. Some freely available links are from sites that do not accept link-checkers at all. The Library of Congress (LC) is an example of this—and the University of Denver has added hundreds of links to content from the LC's THOMAS including legislative histories and other important publications.

When DU had fewer than 10,000 URLs in all, it made sense to use the URL-checking software built into its ILS. But as the numbers continued to grow, and for the reasons outlined above, in 2003 the library discontinued a proactive URL maintenance program in favor of reactive URL maintenance.

Each evening the list of URLs accessed is downloaded into a locally created Microsoft Access database. Each URL is individually verified and corrected if necessary. Obviously in an ideal world it would be optimal if URLs could be corrected before a user has an access error. But since that is not possible, DU opted for the next best situation, and it has worked well.

So what needs to be maintained anyway? In FY2009, just over 55 percent of clickthroughs were to PURLs, but since URL-checking software will not work with PURLs, and since individual libraries should not have to manage GPO's PURLs database, there is no good reason to manage these on the local level. When non-working PURLs are discovered, however, libraries should report these to GPO for the benefit of all catalogs with these records.

Also not needing checking are any commercial links in document records. The University of Denver has rather high utilization of these links, with 667 clickthroughs to LexisNexis content, 241 clickthroughs to Readex content, 114 clickthroughs to HeinOnline content, and 12 other commercial links, accounting for 13.5 percent of total use in FY2009. The remaining 31.5 percent of the clickthroughs are government documents hosted on

.gov, .mil, .org, or any number of other sites that would need to be systematically checked under a proactive link-checking model. However, the difficulty of isolating the many domains makes this too difficult a task. So DU has opted for the reactive approach of fixing URLs after the first errant clickthrough. This approach, though deviating from all examples in the literature, is actually quite effective. If it appears that the clickthrough error is indicative of a server relocation, then all similar links are updated in the OPAC. If a PURL is dead, GPO is immediately notified so that all depositories benefit from the update. If GPO is unable to repair the PURL, substitute content is sought from the Internet Archive, Hathi Trust, or Google Books. If no substitute is found, then the link is withdrawn.

Table 8 shows the University of Denver's error rate over the six-year period.

It appears that our reactive strategy has been able to reduce our error rate significantly over a six-year period.

Conclusion

URLs for U.S. federal documents are becoming available at an ever-increasing rate. Not only is the GPO providing links to nearly all depository content, but retrospective projects, both free and commercial, are adding to the "tonnage." We live in a statistics-driven world, and user accesses to electronic publications are among the more elusive of library statistics. Clickthrough statistics, such as those presented here from the University of Denver, give evidence that users are very much interested in government information. Based on these data, it appears that:

1. DU users prefer online access to documents over print;

2. Users are making use of previously difficult-to-discover resources such as the content within the U.S. Congressional Serial Set;

3. Congressional hearings are the most used parts of the documents collection;

4. Users discover even the most obscure documents resources, and libraries should endeavor to add as much online content as possible to meet users' needs;

5. If online links to older documents are added in the catalog, users will discover and use them;

6. Large data loads, both commercial and freely available, are worth the effort and will get significant use;

7. If a library needs to place documents in storage or withdraw large parts of a collection, there are ways to mitigate the situation by aggressive URL addition projects; and

8. URL tracking technologies can be implemented to track online access and demonstrate the value of documents and documents records in the catalog.

Online access to federal documents places two decisions squarely before library directors and depository librarians: first, how much longer should a library keep receiving documents in tangible formats; and second, what does a library do with the vast amounts of paper documents that are occupying so many linear feet in the stacks? It was as a direct result of the clickthrough statistics presented in this article that DU made the decision to keep virtually all older (legacy) print documents, but at the same time to reduce ongoing receipts of tangible formats to virtually nothing. These decisions simultaneously validate the hundred years of collection development efforts within our documents collection and acknowledge the user preference for online access into the future.

TABLE 8
Error Rates for URL Clickthroughs by Fiscal Year

Fiscal Year	Clicks	Errors	Rate
FY2004	3,809	202	5.30%
FY2005	4,504	231	5.12%
FY2006	4,686	299	6.38%
FY2007	5,217	217	4.15%
FY2008	6,342	179	2.82%
FY2009	7,660	177	2.31%

Appendix: SuDocs Classes

A Department of Agriculture (1862-)
AC Arms Control and Disarmament Agency (1961-)
AE National Archives (1934-)
B Broadcasting Board of Governors/International Broadcasting Bureau
C Department of Commerce (1913-)
CAB Civil Aeronautics Board (1978-1985)
CC Federal Communications Commission (1934-)
CR United States Commission on Civil Rights (1957-)
D Department of Defense (1949-)
E Department of Energy (1977-)
ED Department of Education (1979-)
EP Environmental Protection Agency (1970-)
FCA Farm Credit Administration (1933-1939, 1953-)
FEM Federal Emergency Management Agency (1979-)
FHF Federal Housing Finance Board (1989-)
FM Federal Mediation and Conciliation Service (1947-)
FP Federal Power Commission (1920-1977)
FR Federal Reserve System Board of Governors (1935-)
FS Department of Health, Education and Welfare, etc. (1939-1969)
FT Federal Trade Commission (1914-)
FTZ Foreign-Trade Zones Board (1934-)
FW Federal Works Agency (1939-1949)
GA General Accounting Office (1921-)
GB Geographic Board (1906-1934)
GP Government Printing Office/Superintendent of Documents (1895-)
GS General Services Administration (1949-)
HE Department of Health and Human Services (1980-)
HH Department of Housing and Urban Development (1965-)
HS Department of Homeland Security (2003-)
I Department of the Interior (1849-)
IC Interstate Commerce Commission (1987-1995)
ID U.S. Agency for International Development (1961-)
ITC United States International Trade Commission (1981-)
J Department of Justice (1870-)
JU U.S. Courts (1790-)
L Department of Labor (1913-)
LC Library of Congress (1800-)
LR National Labor Relations Board (1935-)
MS Merit Systems Protection Board (1979-)
N Department of the Navy (1798-1947)
NA National Academy of Sciences (1863-)
NAS National Aeronautics and Space Administration (1958-)
NC National Capital Planning Commission (1952-)
NCU National Credit Union Administration (1970-)
NF National Foundation on the Arts and the Humanities (1965-)
NMB National Mediation Board (1934-)
NS National Science Foundation (1950-)
OP Overseas Private Investment Corporation (1969-)
P United States Postal Service (1970-)

PE Peace Corps (1982-)
PM Office of Personnel Management (1979-)
PR President of the United States (1789-)
PREX Executive Office of the President (1961-)
PRVP Vice President of the United States
S Department of State (1789-)
SBA Small Business Administration (1953-)
SE Securities and Exchange Commission (1934-)
SI Smithsonian Institution (1846-)
SSA Social Security Administration (1995-)
T Department of the Treasury (1789-)
TD Department of Transportation (1966-)
TDA U.S. Trade and Development Agency
VA Department of Veterans Affairs (1930-)
VE Federal Board for Vocational Education (1917-1933)
W Department of War (1789-1947)
X Congress
Y 1 Congress
Y 3 Congress: Commissions, Committees, Boards
Y 4 Congress: Hearings, Committee Prints, Publications
Y 7 Memorial Addresses
Y 6 Impeachment Proceedings
Y 10 Congressional Budget Office
Y 11 Congressional Office of Compliance
Z Continental Congress Papers

Notes

1. Christopher Brown, "Knowing Where They're Going: Statistics for Online Government Document Access through the OPAC," *Online Information Review* 28, no. 6 (2004): 396–409.

2. The law that established GPO Access was Public Law 103-40, the Government Printing Office Electronic Information Access Enhancement Act of 1993. See also United States General Accounting Office, *Information Management: Electronic Dissemination of Government Publications* (Washington, D.C.: GAO, 2001), 54. Available online at www.gao.gov/new.items/d01428.pdf. [Accessed 10 January 2010].

3 U.S. Government Printing Office, *Annual Report Fiscal Year 2002*. Available online at www.gpo.gov/pdfs/congressional/archives/2002gpoannualreport.pdf, 7. The modern Federal Depository Library Program began with the Depository Library Act of 1962 (Public Law 87-579), which initiated the system of regional depository libraries and selective depository libraries; it was the basis for all the strict regulations that emanated from the Act. In 1993, the Government Printing Office Electronic Information Access Enhancement Act of 1993 (Public Law 103-40) was the legislation that effectively launched the electronic era of documents with the establishment of GPO Access (www.gpoaccess.gov).

4. Ric Davis, *Federal Depository Library Conference and Fall Depository Library Council Meeting Proceedings, Oct. 20, 2008,* available online at www.fdlp.gov/home/repository/doc_view/193-transcript-of-proceedings-november-20-2008 [accessed June 6, 2009]. See also Depository Library Council, Federal Depository Library Program Strategic Plan, 2009–2014, Draft Discussion Document, Apr. 17, 2009, available online at www.fdlp.gov/home/repository/doc_view/649-federal-depository-library-program-strategic-plan-2009–2014 [accessed 6 June 2009].

5. Peter Hernon and Laura Saunders, "The Federal Depository Library Program in 2023: One Perspective on the Transition to the Future," *College & Research Libraries* 70, no. 4 (2009): 351–70.

6. Documents without Shelves. Available online at www.marcive.com/HOMEPAGE/docs-wout.htm. [Accessed 6 June 2009].

7. Thomas E. Nisonger, "A Simulated Electronic Availability Study of Serial Articles through a University Library Web Page," *College & Research Libraries* 70, no. 5 (2009): 422–45.

    8. Michael Levine-Clark, "Electronic Book Usage: A Survey at the University of Denver," *Portal (Baltimore, Md.)* 6, no. 3 (2006): 285.

    9. Aparna Zambare et al., "Assuring Access: One Library's Journey from Print to Electronic Only Subscriptions," *Serials Review* 35, no. 2 (2009): 70–74.

    10. Jennifer Gerke and Jack M. Maness, "The Physical and the Virtual: The Relationship between Library as Place and Electronic Collections," *College & Research Libraries* 71, no. 1 (2010): 20–31.

    11. Mary Schneider Laskowski, "The Impact of Electronic Access to Government Information: What Users and Documents Specialists Think," *Journal of Government Information* 27, no. 2 (2000): 173–85.

    12. Jennie M. Burroughs, "What Users Want: Assessing Government Information Preferences to Drive Information Services," *Government Information Quarterly* 26, no. 1 (2009): 203–18.

    13. Roger C. Schonfeld and Ross Housewright, *Documents for a Digital Democracy: A Model for the Federal Depository Library Program in the 21st Century* (2009). Available online at www.ithaka.org/ithaka-s-r/research/documents-for-a-digital-democracy/Documents%20for%20a%20Digital%20Democracy.pdf. [Accessed 10 January 2010].

    14. Brown, "Knowing Where They're Going."

    15. Susan L. Kendall et al., "Tracking Usage of E-government Publications," *The Serials Librarian* 50, no. 3/4 (2006): 295–303.

    16. Susan Kendall, pers. comm., Jan. 12, 2010.

    17. Stephanie Braunstein, pers. comm., June 29, 2009 and Aug. 7, 2009.

    18. Liza Weisbrod, pers. comm., June 17, 2009 and Aug. 8, 2009.

    19. Jennifer Gerke, pers. comm., Nov. 15, 2010

    20. Shane Cole, pers. comm., June 30, 2009.

    21. Christopher C. Brown, "Tracking Online Document Usage from the Catalog: Experiences from the Field," presentation given with Stephanie Braunstein, Susan Kendall, Liza Weisbrod, Jennifer Gerke, and Shane Cole at the Fall 2009 Depository Library Conference, Oct. 19, 2009, Arlington, Va. Available online at www.fdlp.gov/home/repository/doc_download/1051-tracking-online-document-usage-from-the-catalog. [Accessed 10 January 2010].

    22. Liya Deng, "Staying or Leaving? Why We Don't Have to Make the Choice," presentation given at the Fall Depository Library Conference, Oct. 21, 2009, Arlington Va. Available online at www.fdlp.gov/home/repository/doc_download/1055-fdlp-staying-or-leaving-why-we-dont-have-to-make-the-choice. [Accessed 19 January 2010].

    23. Federal Depository Library PURL Referrals. Available online at www.gpo.gov/su_docs/fdlp/coll-dev/referrals/. [Accessed 10 January 2010].

    24. Burroughs, "What Users Want," 204.

    25. Gerald Burke et al., "URLs in the OPAC: Integrating or Disintegrating Research Libraries' Catalogs," *Journal of Academic Librarianship* 29, no. 5 (2007): 290–97.

    26. Brown, "Knowing Where They're Going," 398.

    27. Nicholas N. Joint, "URLs in the OPAC: Comparative Reflections on US vs UK Practice," *Library Review (Glasgow)* 56, no. 3 (2007): 182–93.

    28. Ibid., 188.

    29. Ibid., 188.

    30. C. Rockelle Strader and Farrell D. Hamill, "Rotten but not Forgotten: Weeding and Maintenance of URLs for Electronic Resources in The Ohio State University Online Catalog," *Serials Librarian* 53, no. 1/2 (2007): 163–77.

    31. Christopher C. Brown and Elizabeth S. Meagher, "Cataloging Free E-resources: Is It Worth the Investment?" *Interlending & Document Supply* 36, no. 3 (2008): 135–41.

    32. Elizabeth S. Meagher and Christopher C. Brown, "Turned Loose in the OPAC: URL Selection, Addition, and Management Process," *Library Hi Tech* 28, no. 3 (2010): 360–76.

    33. Terraserver. Available online at www.terraserver.com/. [Accessed 10 January 2010].: Recently, however, Microsoft migrated their OpenGIS Map Server from terraserver.com to msrmaps.com [Accessed November 15, 2010].

    34. Christopher C. Brown, "Topos in Your OPAC: Linking to the Online Version," *Information Bulletin* / Western Association of Map Libraries 34, no. 3 (2003): 141–44.

    35. The General Accounting Office was established in 1921, but the name was changed to the Government Accountability Office in 2004. The acronym GAO is used to refer to both names. The bulk of the records for this URL-addition project span the years 1976–2000.

    36. Brown, "Knowing Where They're Going"; Christopher C. Brown, "Statistics for Online Document Use," presentation given at the Fall 2003 Depository Library Conference, Oct. 22, 2003, Arlington, Va., published in the *Proceedings of the 12th Annual Depository Library Conference, Oct. 19–22, 2003*, available online at http://www.fdlp.gov/home/repository/doc_download/710-statistics-for-online-document-use [accessed 10 January 2010]; Christopher C. Brown, "Knowing

Where They're Going: Counting Field 856 Click-throughs in the OPAC," presentation given at the 12th Annual Innovative Users Group, Aprl 3 & 5, 2004, Boston, Mass.; Christopher C. Brown, "Tracking Online Document Usage from the Catalog: Experiences from the Field," presentation given at the Fall Depository Library Conference, Oct. 19 2009, Arlington Va., available online at www.fdlp.gov/home/repository/doc_download/1051-tracking-online-document-usage-from-the-catalog [accessed 10 January 2010].

37. Barbara J. Costello, "A Snapshot of Availability of U.S. Congressional Committee Hearings: How Complete Are Online Sources?" *College & Undergraduate Libraries* 10, no. 2 (2003): 73–87.

38. *FDLP Handbook*, Chapter 12, www.fdlp.gov/administration/handbook/127?start=14. [Accessed 19 January 2010].

39. Christopher C. Brown, "Adding URLs in Bulk at the University of Denver," presentation given at the Spring 2002 Depository Library Council Meeting, Apr. 24, 2002, Mobile, Ala. Available online at www.fdlp.gov/home/repository/doc_download/512-urls-and-purls-in-online-public-access-catalogs-a-web-pages. [Accessed 19 January 2010].

40. Christopher C. Brown, "Topos in Your OPAC: Linking to the Online Version," presentation given at the Western Association of Map Libraries Spring 2003 conference, Mar. 28, 2003, Stanford University, Calif., http://sites.google.com/site/librariancorner/powerpoints; Christopher C. Brown, "Topos in Your OPAC: Linking to the Online Version," *Information Bulletin* / Western Association of Map Libraries 34, no. 3 (2003): 141–44.

41. Christopher C. Brown, "Enhancing NASA Fiche Records with Links to Online Content," presentation given at the Fall 2007 Depository Library Conference, Oct. 17, 2007, Arlington, Va. Available online at www.fdlp.gov/component/docman/doc_download/283-enhancing-nasa-fiche-records-with-links-to-online-content. [Accessed 10 January 2010].