SEARCH ENGINE OPTIMIZATION FOR DIGITAL COLLECTIONS

Kenning Arlitsch
Patrick OBrien
Sandra McIntyre

Agenda

- □ Assessment
- □ Phase 1: Start feedback loop
- □ Phase 2: Get indexed
- □ Phase 3: Increase visibility (future)
- □ Wrap-up

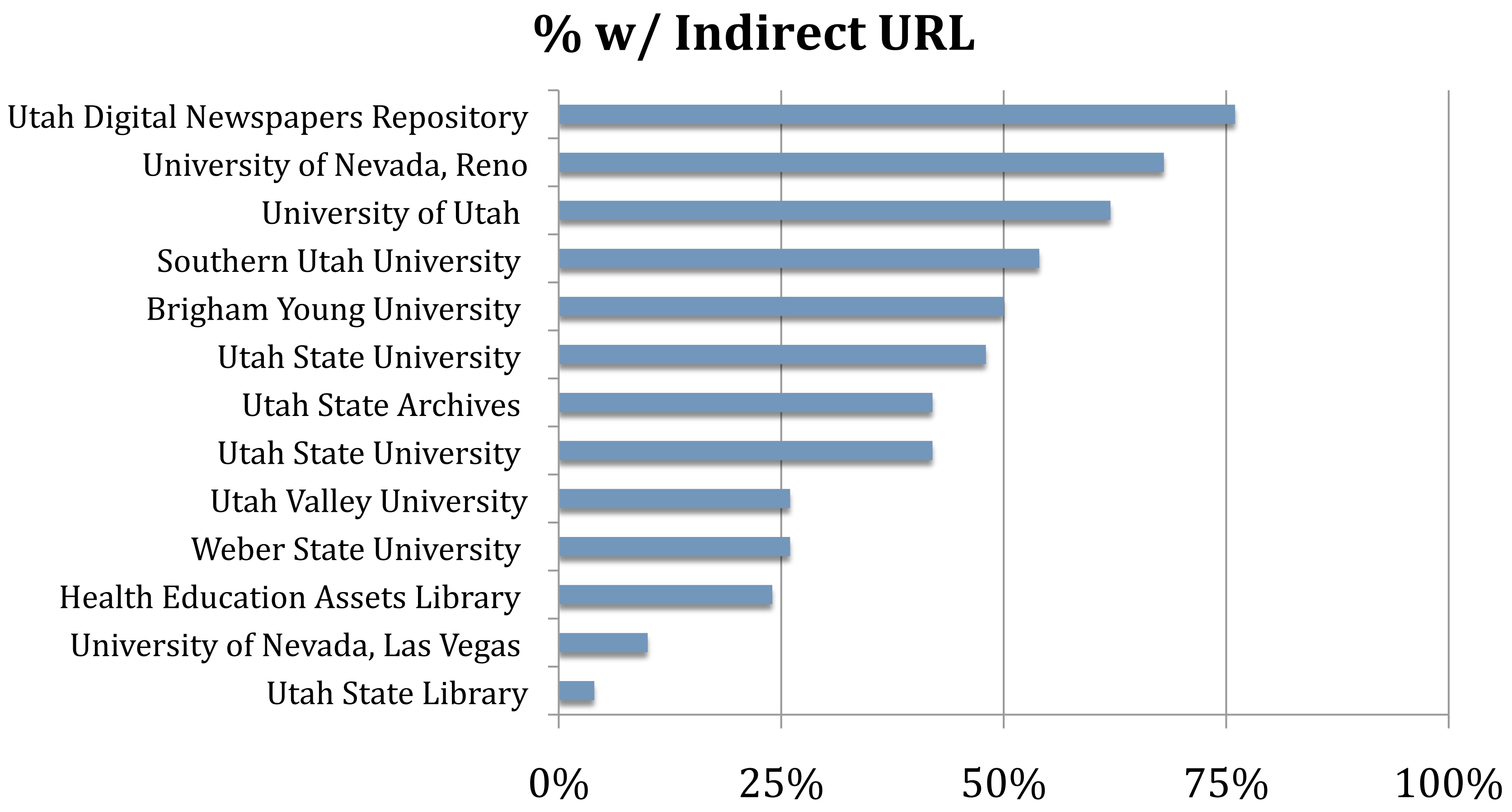
Context and history at Utah

- Large digital library programs
  - Mountain West Digital Library
  - Utah Digital Newspapers
  - Western Soundscape Archive
  - Western Waters Digital Library
- Digital collections are "Deep Web"
- Google indexing diminished recently
  - Ceased OAI harvest in August 2008
  - Average as low as 8% in spring 2010

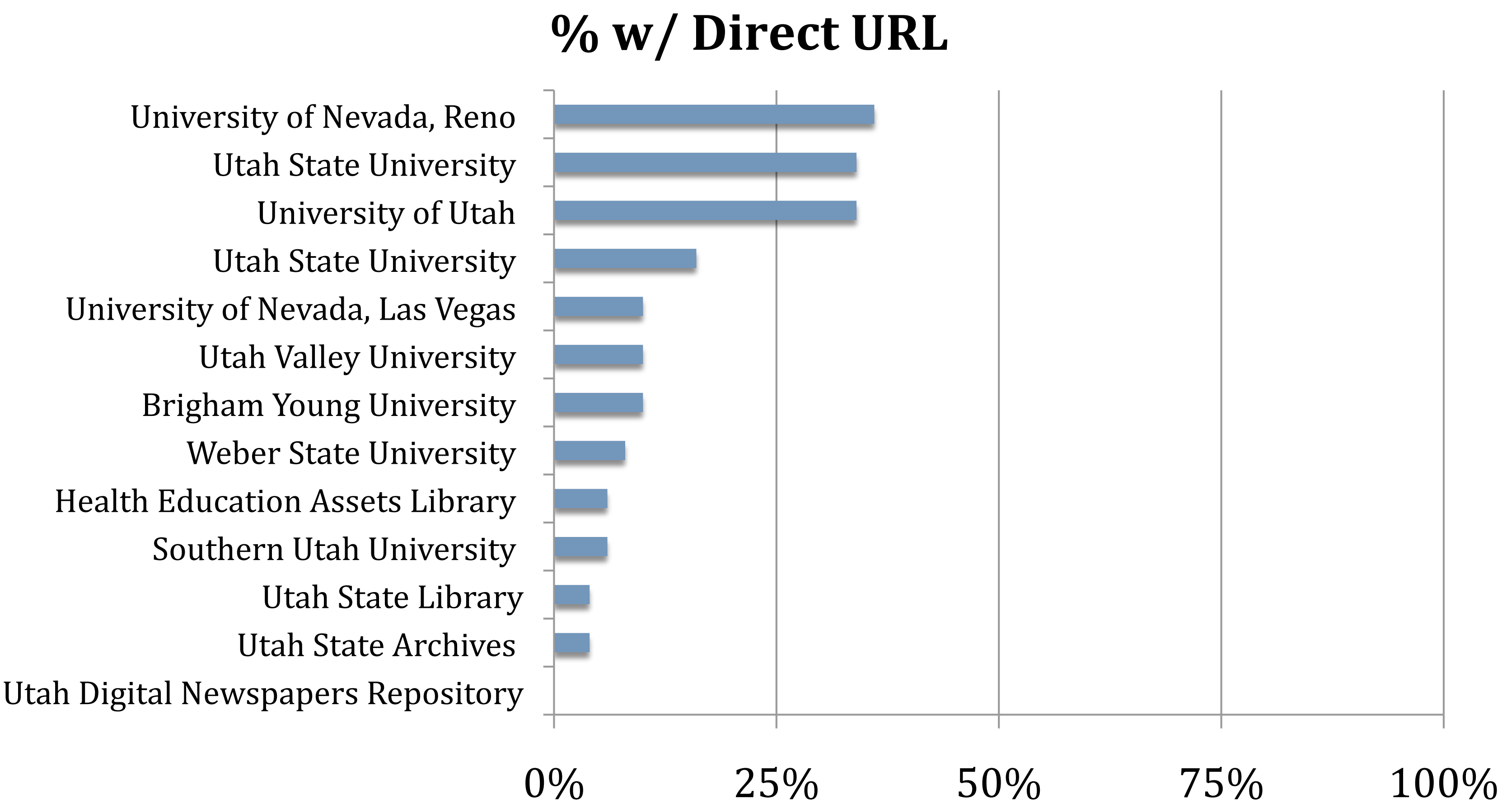
Initial Repositories Survey

- Surveyed 13 repositories of the MWDL in July
  - 10 CONTENTdm
  - 1 Digital Commons
  - 1 ArchivalWare
  - 1 home grown (HEAL)
- Randomly selected 50 objects from each (650)
- Searched by title in Google and Google Images
  - 38% find rate in Google
  - Almost 0% in Google Images

MWDL Repositories Survey



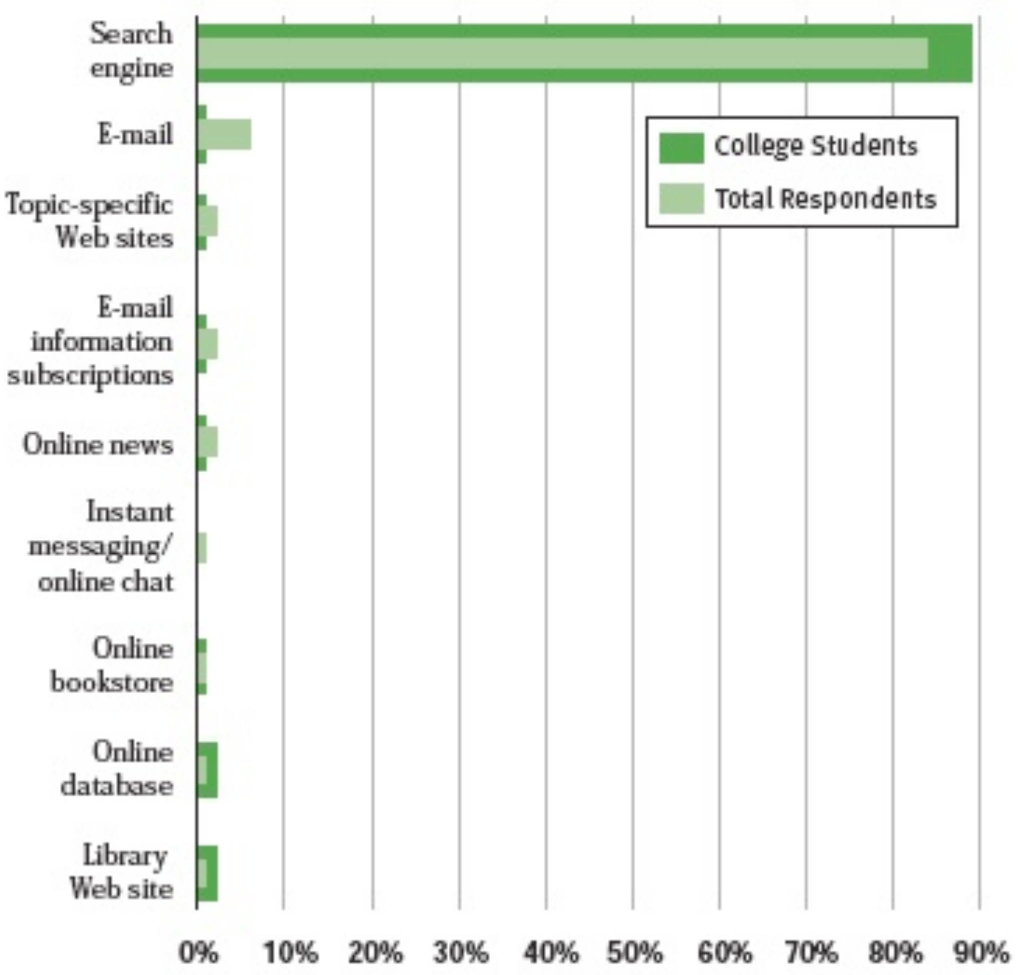
MWDL Repositories Survey



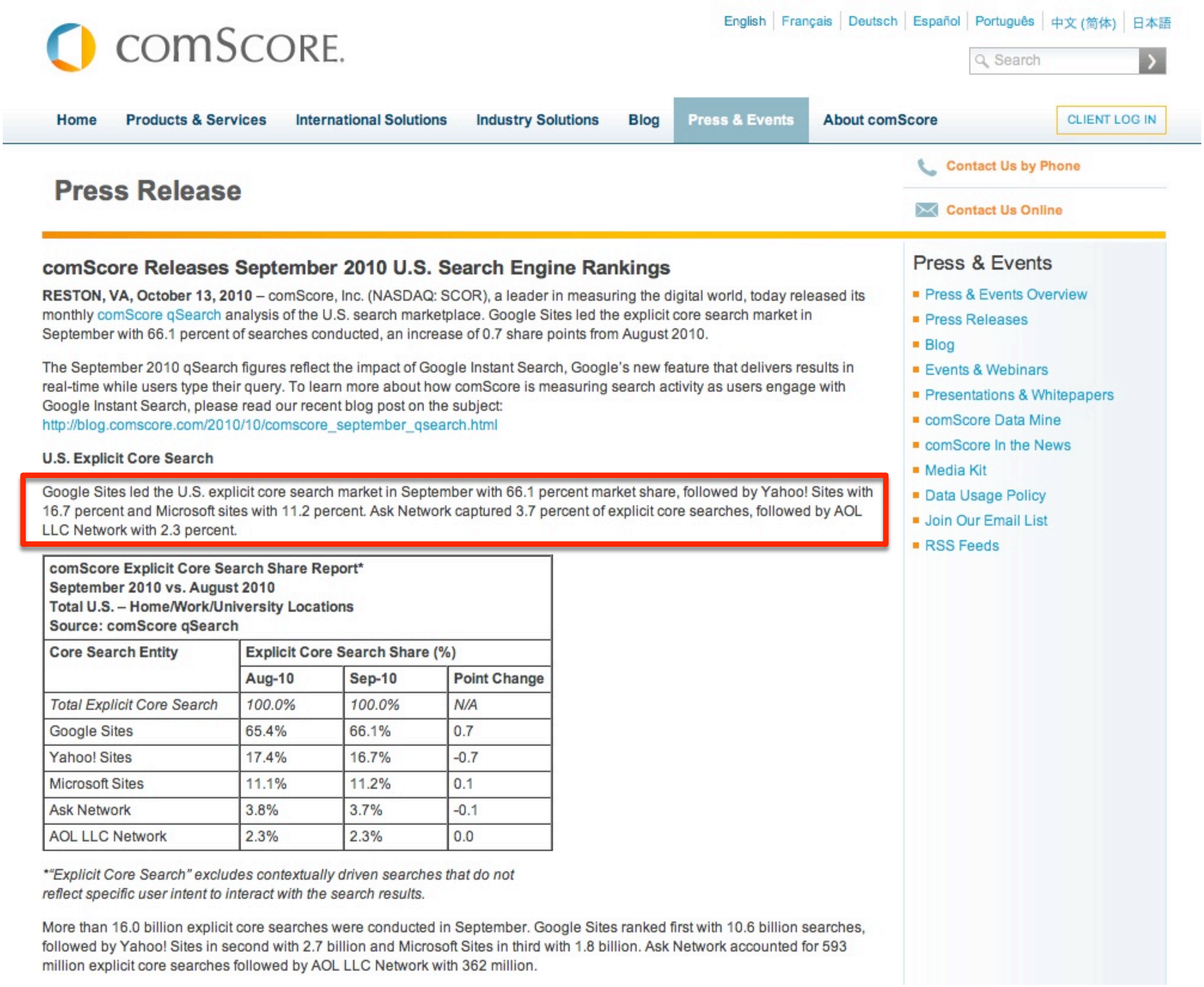
Discoverability of digital resources

- Priority Collections
  - Institutional Repository (USpace)
  - Special Collections EAD finding aids
  - University Press
- Discoverability is important for
  - Faculty (contributors and users)
  - Donors
  - Students

Where College Students Begin Searching



Source: *Perceptions of Libraries and Information Resources*, OCLC, 2005, question 520.
Note: Only electronic resources with usage rates of 1 percent or more are represented on this graph.



Press Release

comScore Releases September 2010 U.S. Search Engine Rankings

RESTON, VA, October 13, 2010 – comScore, Inc. (NASDAQ: SCOR), a leader in measuring the digital world, today released its monthly comScore qSearch analysis of the U.S. search marketplace. Google Sites led the explicit core search market in September with 66.1 percent of searches conducted, an increase of 0.7 share points from August 2010.

The September 2010 qSearch figures reflect the impact of Google Instant Search, Google's new feature that delivers results in real-time while users type their query. To learn more about how comScore is measuring search activity as users engage with Google Instant Search, please read our recent blog post on the subject: http://blog.comscore.com/2010/10/comscore_september_qsearch.html

U.S. Explicit Core Search

Google Sites led the U.S. explicit core search market in September with 66.1 percent market share, followed by Yahoo! Sites with 16.7 percent and Microsoft sites with 11.2 percent. Ask Network captured 3.7 percent of explicit core searches, followed by AOL LLC Network with 2.3 percent.

comScore Explicit Core Search Share Report*
September 2010 vs. August 2010
Total U.S. – Home/Work/University Locations
Source: comScore qSearch

Core Search Entity	Explicit Core Search Share (%)		
	Aug-10	Sep-10	Point Change
Total Explicit Core Search	*100.0%*	*100.0%*	*N/A*
Google Sites	65.4%	66.1%	0.7
Yahoo! Sites	17.4%	16.7%	-0.7
Microsoft Sites	11.1%	11.2%	0.1
Ask Network	3.8%	3.7%	-0.1
AOL LLC Network	2.3%	2.3%	0.0

**"Explicit Core Search" excludes contextually driven searches that do not reflect specific user intent to interact with the search results.*

More than 16.0 billion explicit core searches were conducted in September. Google Sites ranked first with 10.6 billion searches, followed by Yahoo! Sites in second with 2.7 billion and Microsoft Sites in third with 1.8 billion. Ask Network accounted for 593 million explicit core searches followed by AOL LLC Network with 362 million.

Literature Review

- Googlizing a Digital Library. By: DeRidder, Jody L. ,Code4Lib Journal, 2008.
- Worst Practices in Search Engine Optimization. MALAGA, ROSS A.. Communications of the ACM, Dec2008, Vol. 51 Issue 12, p147-150
- Searching for a New Way to Reach Patrons: A Search Engine Optimization Pilot Project at Binghamton University Libraries. By: Rushton, Erin E.; Kelehan, Martha Daisy; Strong, Marcy A.. Journal of Web Librarianship, 2008, Vol. 2 Issue 4, p525-547
- Optimal Results: What Libraries Need to Know About Google and Search Engine Optimization. By: Cahill, Kay; Chalut, Renee. Reference Librarian, Jul-Sep2009, Vol. 50 Issue 3, p234-247
- Academic Search Engine Optimization. By: Beel, Jöran; Gipp, Bela; Eilde, Erik. Journal of Scholarly Publishing, Jan2010, Vol. 41 Issue 2, p176-190

Literature Lessons

- Most are dated
- Most deal with general websites
- “Black hat” techniques get you banned
- Few deal with digital collections in db’s
- Some suggest duplicating the content outside the database

Problems evident on several levels

- Web server
  - robots.txt
  - Crawler errors
- Application layer (repository software)
  - URL redirects
  - Many URLs for same objects
- Presentation layer
  - HTML and Graphic design
- Metadata issues

External Influence: Search Engine Policies

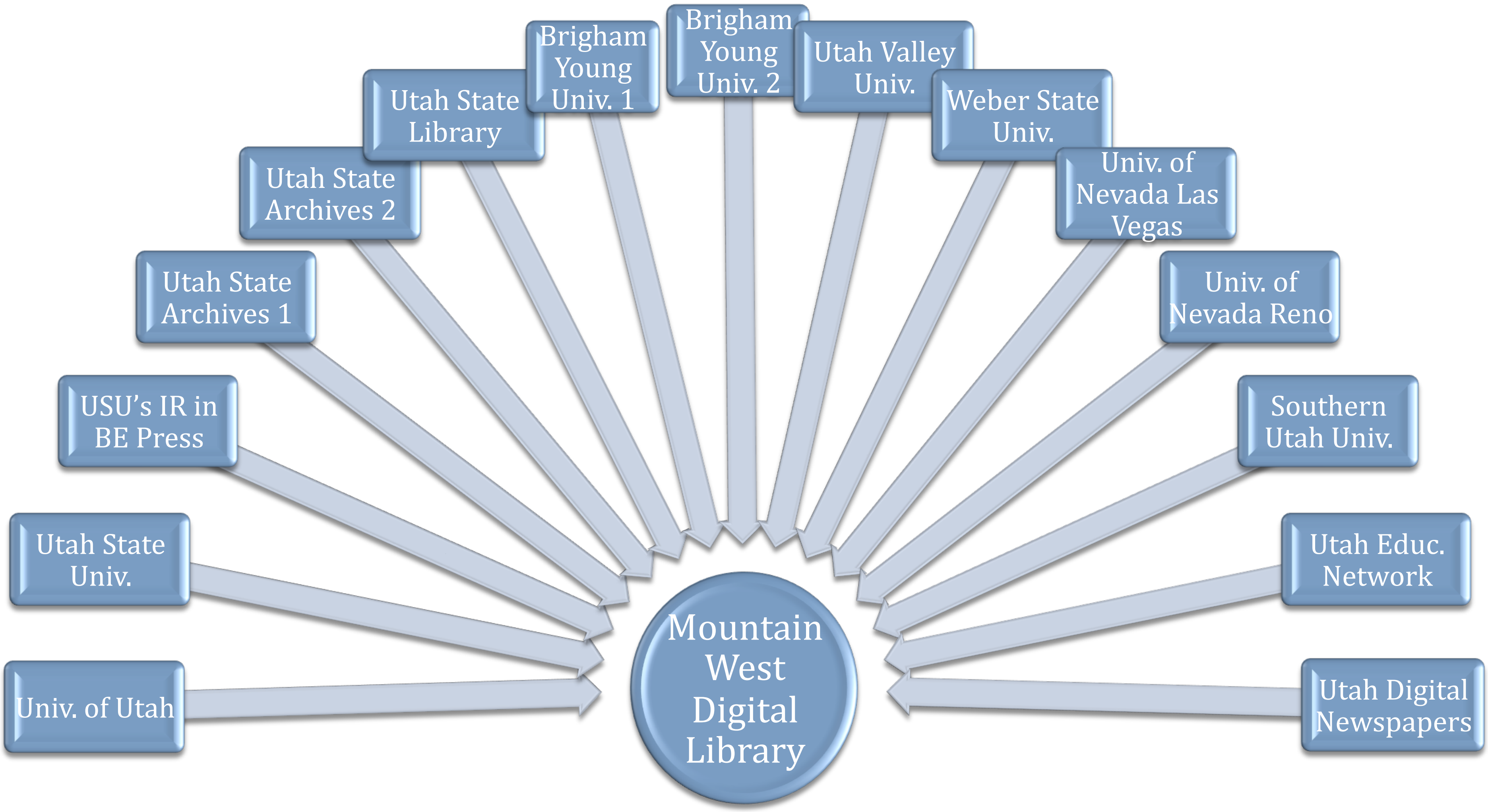
- Rules and enforcement levels change
  - OAI harvesting
  - Sitemaps
- Requirements & standards adoption
  - W3C, Highwire, etc.
- Insensitive to standards valued by librarians
  - "Use Dublin Core tags (e.g., DC.Title) as a last resort"*

* Google Scholar Inclusion Guidelines for Webmasters
http://scholar.google.com/intl/en/scholar/inclusion.html

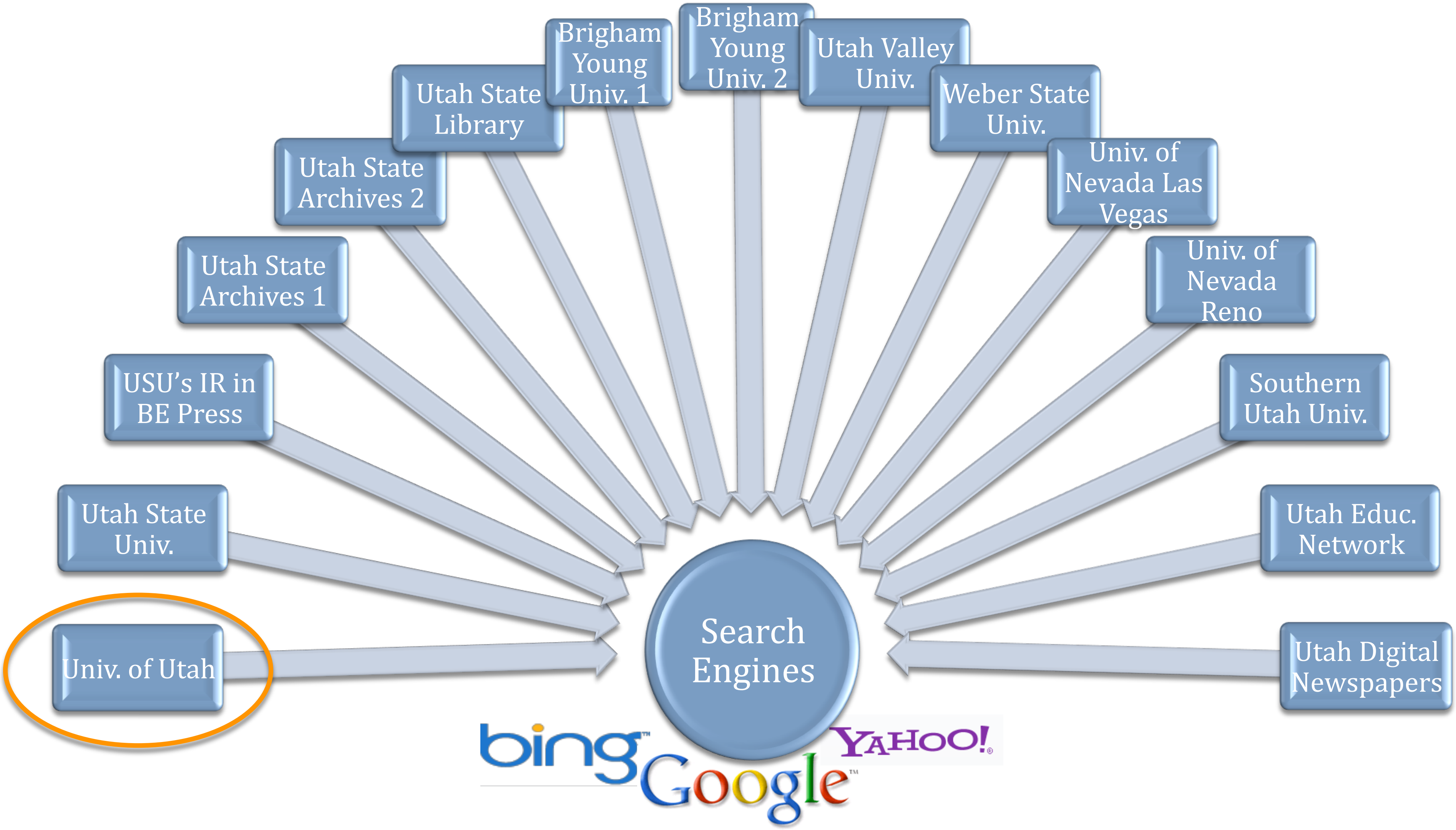
Agenda

- □ Assessment
- □ Phase 1: Start feedback loop
- □ Phase 2: Get indexed
- □ Phase 3: Increase visibility (future)
- □ Wrap-up

Mountain West Digital Library



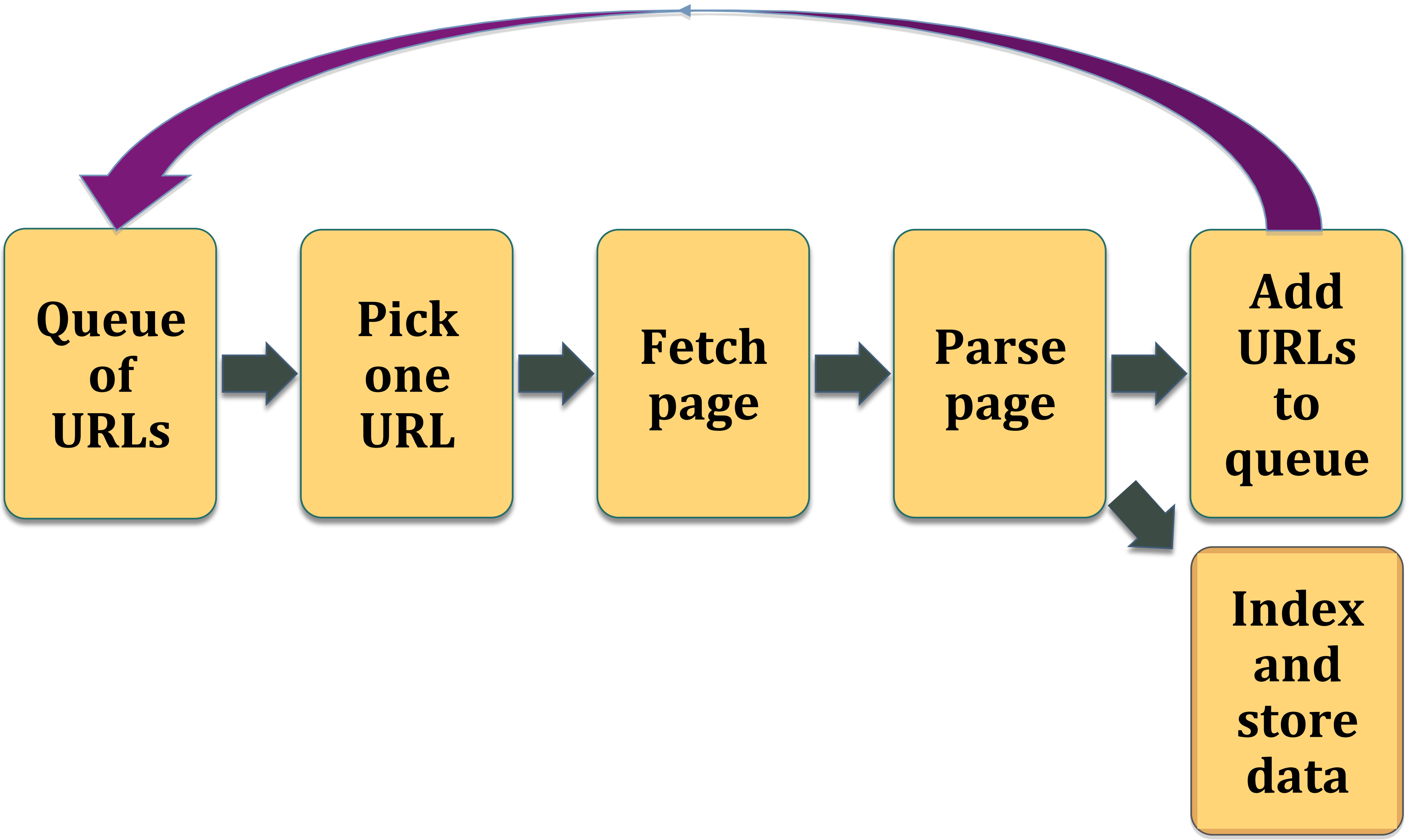
Mountain West Digital Library



Google and Digital Assets Management

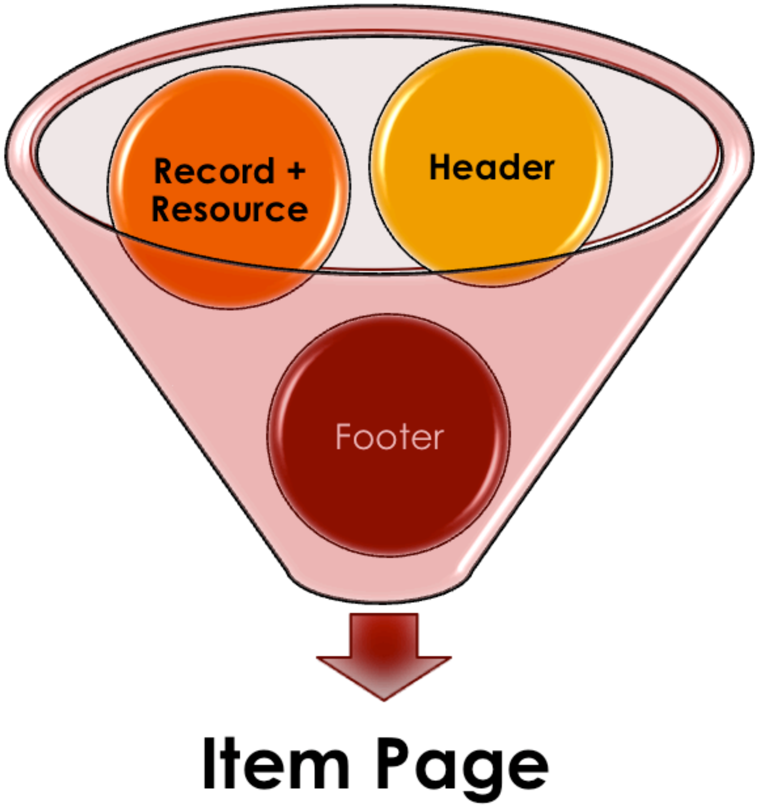
- 2008: Google announced it would no longer crawl Open Archives Initiative (OAI) streams
- Many digital collections have been slowly “disappearing” from Google since then
- What’s going on?
- What’s needed instead?

Phase 1:
Learning about Web Crawlers



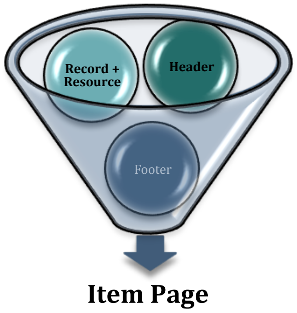
Phase 1:
Notifying crawlers about dynamic pages

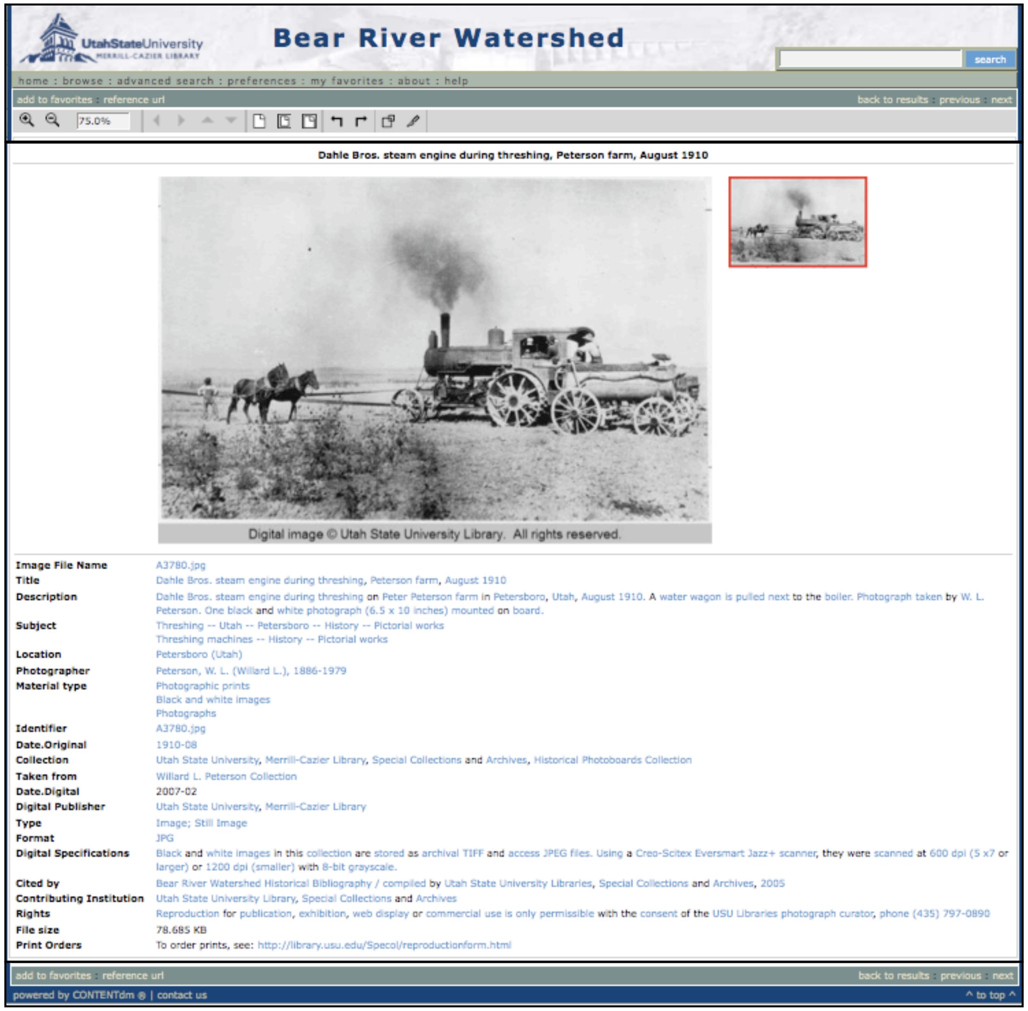
- Digital asset management systems construct pages in HTML on the fly
  - Header
  - Record retrieved from database and formatted
  - Footer



Phase 1:
Notifying crawlers about dynamic pages

- □ Have to tell crawler how to assemble it (with URL)




UtahStateUniversity
MERRILL-CAZIER LIBRARY
search
home : browse : advanced search : preferences : my favorites : about : help
add to favorites : reference url
back to results : previous : next
75.0%

Bear River Watershed

Dahle Bros. steam engine during threshing, Peterson farm, August 1910

Digital image © Utah State University Library. All rights reserved.

Image File Name	A3780.jpg
Title	Dahle Bros. steam engine during threshing, Peterson farm, August 1910
Description	Dahle Bros. steam engine during threshing on Peter Peterson farm in Petersboro, Utah, August 1910. A water wagon is pulled next to the boiler. Photograph taken by W. L. Peterson. One black and white photograph (6.5 x 10 inches) mounted on board.
Subject	Threshing -- Utah -- Petersboro -- History -- Pictorial works Threshing machines -- History -- Pictorial works
Location	Petersboro (Utah)
Photographer	Peterson, W. L. (Willard L.), 1886-1979
Material type	Photographic prints Black and white images Photographs
Identifier	A3780.jpg
Date.Original	1910-08
Collection	Utah State University, Merrill-Cazier Library, Special Collections and Archives, Historical Photoboards Collection
Taken from	Willard L. Peterson Collection
Date.Digital	2007-02
Digital Publisher	Utah State University, Merrill-Cazier Library
Type	Image; Still Image
Format	JPG
Digital Specifications	Black and white images in this collection are stored as archival TIFF and access JPEG files. Using a Creo-Scitex Eversmart Jazz+ scanner, they were scanned at 600 dpi (5 x7 or larger) or 1200 dpi (smaller) with 8-bit grayscale.
Cited by	Bear River Watershed Historical Bibliography / compiled by Utah State University Libraries, Special Collections and Archives, 2005
Contributing Institution	Utah State University Library, Special Collections and Archives
Rights	Reproduction for publication, exhibition, web display or commercial use is only permissible with the consent of the USU Libraries photograph curator, phone (435) 797-0890
File size	78.685 KB
Print Orders	To order prints, see: http://library.usu.edu/Specol/reproductionform.html

add to favorites : reference url
back to results : previous : next
powered by CONTENTdm ® | contact us
^ to top ^

Google Sitemaps

- Sitemap file for each collection

> **“Here is a list of the URLs of the dynamic pages that I want you to crawl, one for each item.”**

- Sitemap Index file to list all the Sitemaps

> **“Here is a list of all the Sitemap files.”**

- Protocol: http://www.sitemaps.org

Start the feedback loop

- Create Sitemaps, one for each collection, and Sitemap Index.
- Register with Google Webmaster Tools.
- Inform Google about the location of Sitemap Index.
  - In Webmaster Tools: http://www.google.com/webmasters/
  - In the robots.txt file at the root on the server
- Monitor crawler results in Webmaster Tools.

Initial experiments and theories: Presentation layer

- □ Compound objects – frameset
- □ Page titles
- □ Putting metadata up in head as <meta> tags

Monitor crawler results

- Webmaster Tools
  - Top search queries
  - Links to your site
  - Keywords
  - Internal links
  - Crawl errors
  - Crawl stats
  - HTML suggestions

Phase 1 results:
Feedback loop is in place

- Webmaster Tools shows us results
  - Incomplete indexing
  - Lots of crawler errors
  - Inconsistencies across collections
  - Low ranking on search engine listings

Cross-departmental collaboration

- Search Engine Optimization (SEO) Team
  - Associate Director for IT Services
    - Server administrators
    - Programmers
    - Digital Initiatives Librarian
  - Collection managers and other metadata experts

- SEO consultant volunteered services:
  Patrick OBrien of RevX Corp.

Agenda

□ Assessment

□ Phase 1: Start feedback loop

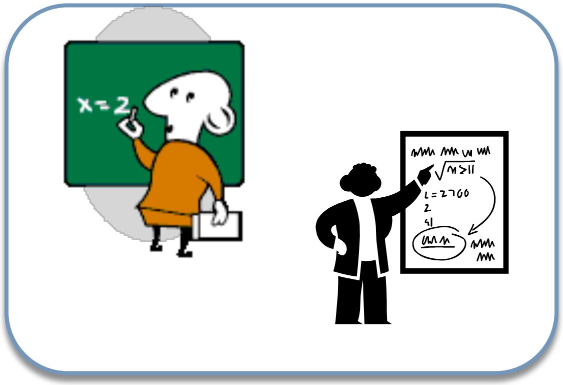
□ Phase 2: Get indexed

□ Phase 3: Increase visibility (future)

□ Wrap-up

Know your customers and what they value.

Faculty



Value

High

- Publication Page Views
- Publication Downloads
- Requests for Information
- Publication Citations

Collection Donors

Value

High

- Digital Collection Pages Indexed
- Digital Collection Page Views
- Digital Collection Visitors
- Requests for More Info
- Physical Collection Visitors
- Reproductions Ordered

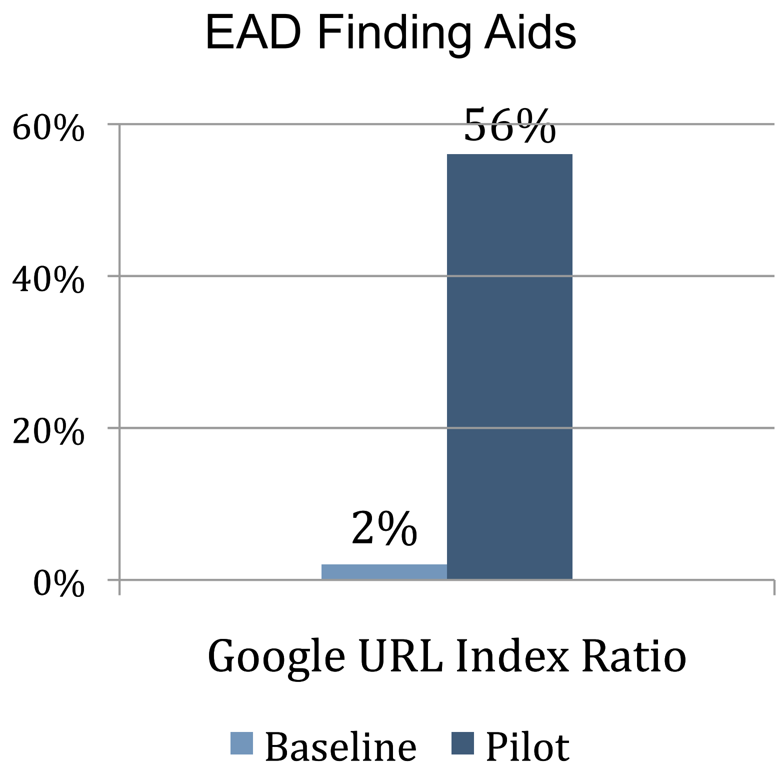
Phase 2 goals and results

Goals

- □ Increase the number of Digital Collection web pages in the Google search engine.
- □ Develop a program to maximize a collections visibility and reach

Pilots

Results



75 pages indexed / 3,212 pages submitted as of April 24, 2010
7,304 pages indexed / 13,035 pages submitted as of November 01, 2010

Phase 2 goals and results

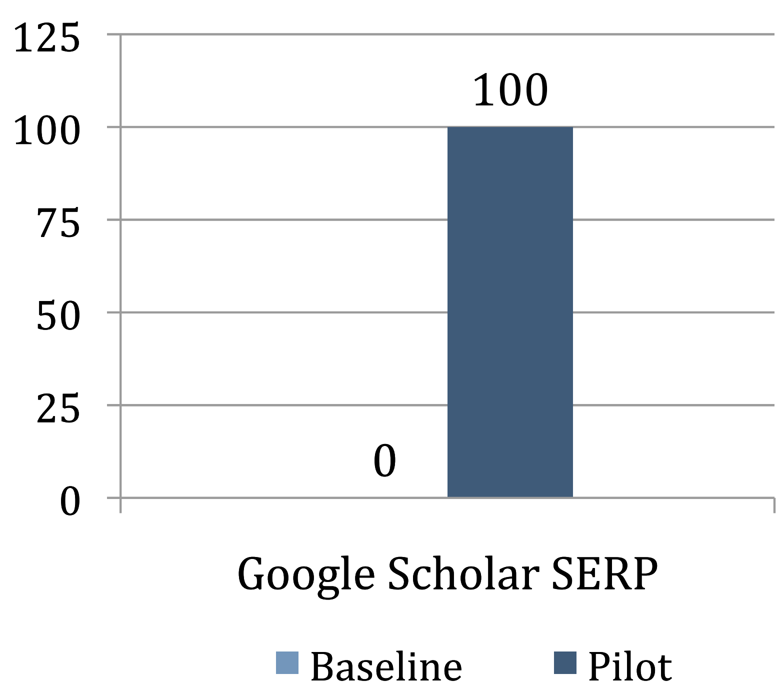
Goals

- □ Increase the number of Digital Collection web pages in the Google search engine.
- □ Develop a program to maximize a collections visibility and reach

Pilots

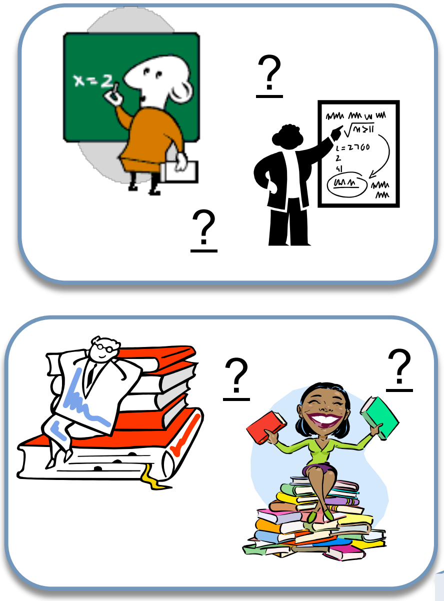
Results

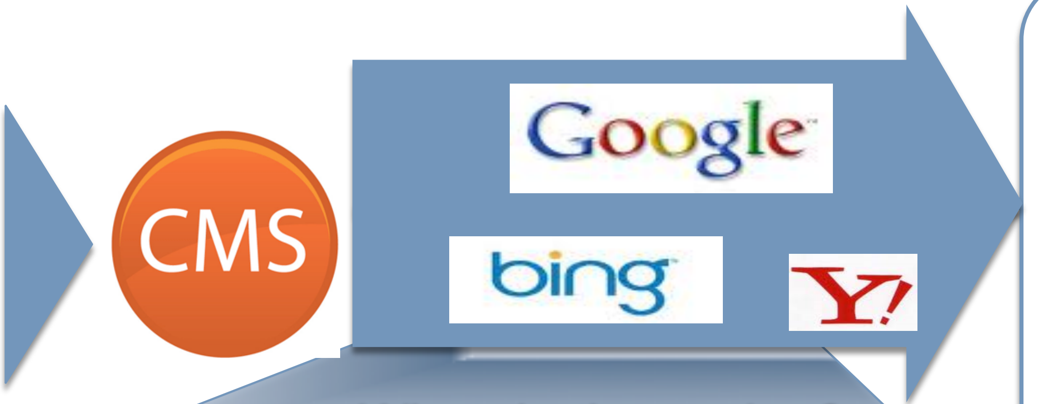
IR Articles*

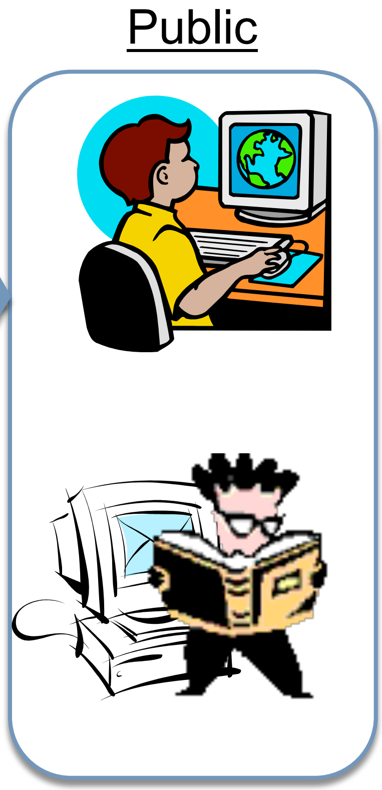


* site:content.lib.utah.edu as of April 24, 2010 November 10, 2010

Why can’t the public find our content?





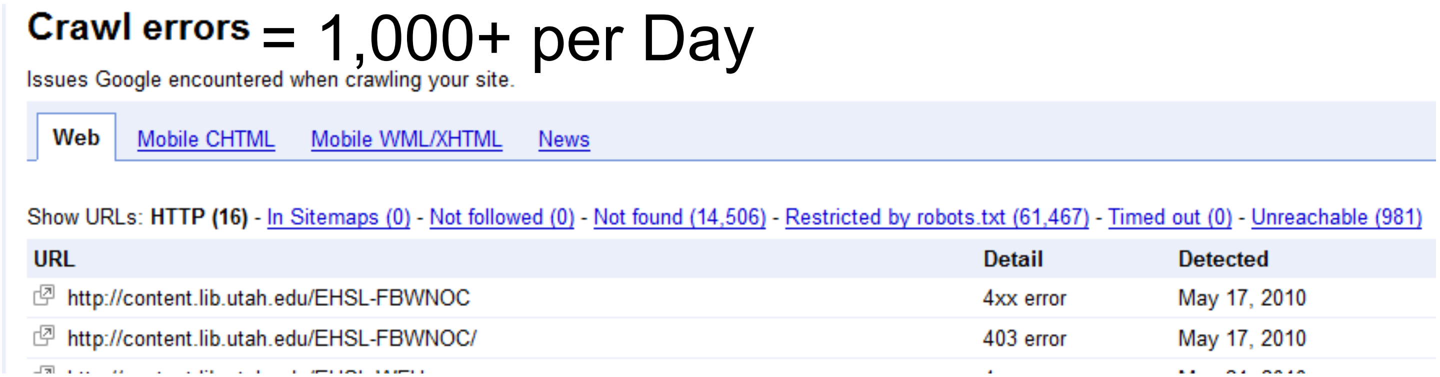


What do they value?

- □ Are you worthy enough for their customer (i.e Index)?
- □ How much will their customer value the introduction (i.e, Visibility)?

The Digital Collection environment is complex and very difficult for robots to index.

- Multiple Web Server Technologies
- Complex Application Platforms
- Different Metadata Organization, Context, and process
- Constantly changing Search Engine Requirements



Crawl errors = 1,000+ per Day

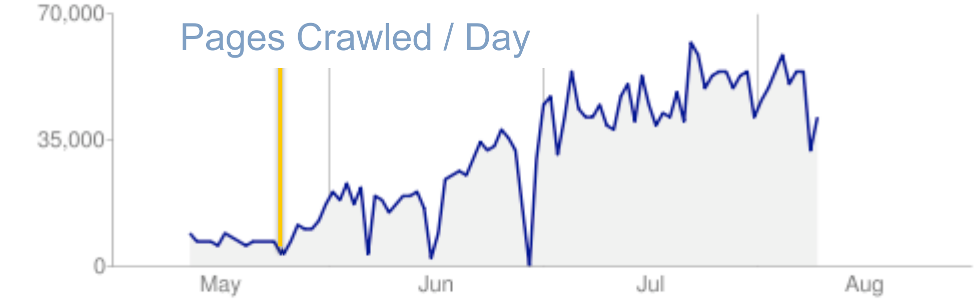
Issues Google encountered when crawling your site.

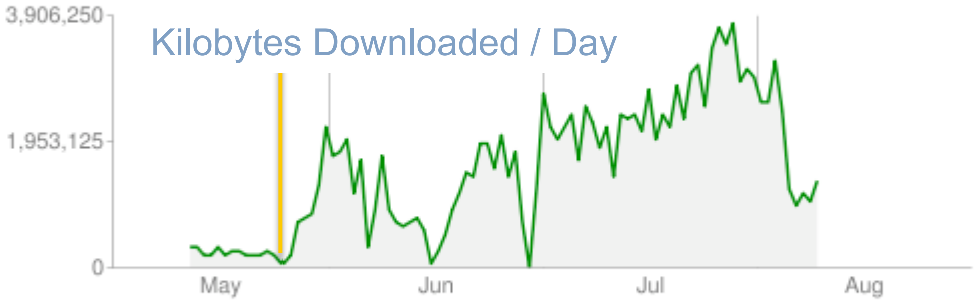
Show URLs: **HTTP (16)** - In Sitemaps (0) - Not followed (0) - Not found (14,506) - Restricted by robots.txt (61,467) - Timed out (0) - Unreachable (981)

URL	Detail	Detected
http://content.lib.utah.edu/EHSL-FBWNOC	4xx error	May 17, 2010
http://content.lib.utah.edu/EHSL-FBWNOC/	403 error	May 17, 2010

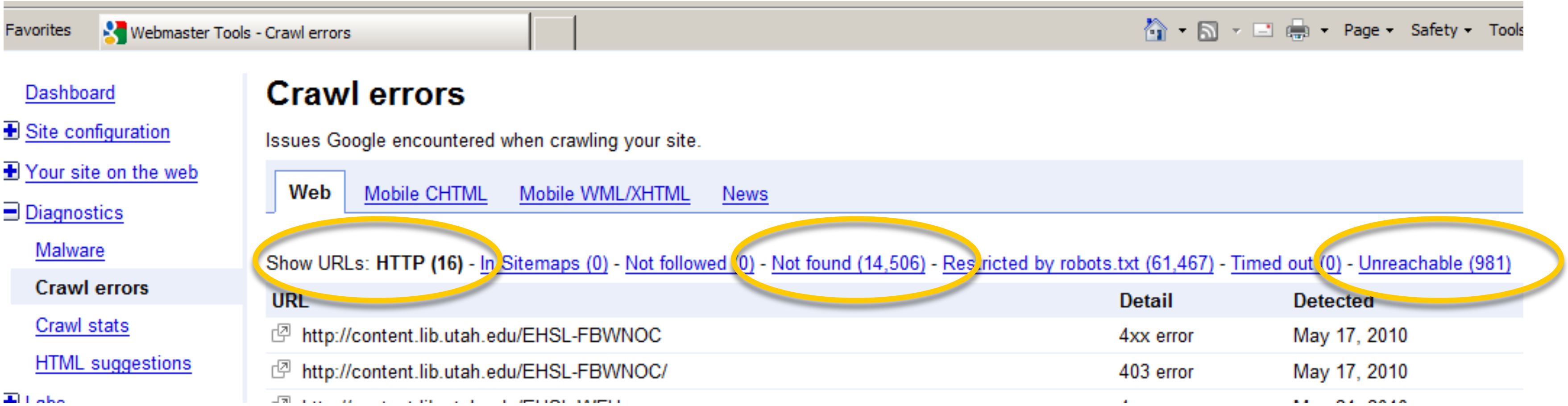
Are you worthy enough for their customers (i.e Index)

- Reduce Google Crawl Errors
- Developed efficient Google Crawler path
- Reconfigured the environment to meet Google's requirements





Check the Crawl Errors in Google Webmaster



Crawl errors

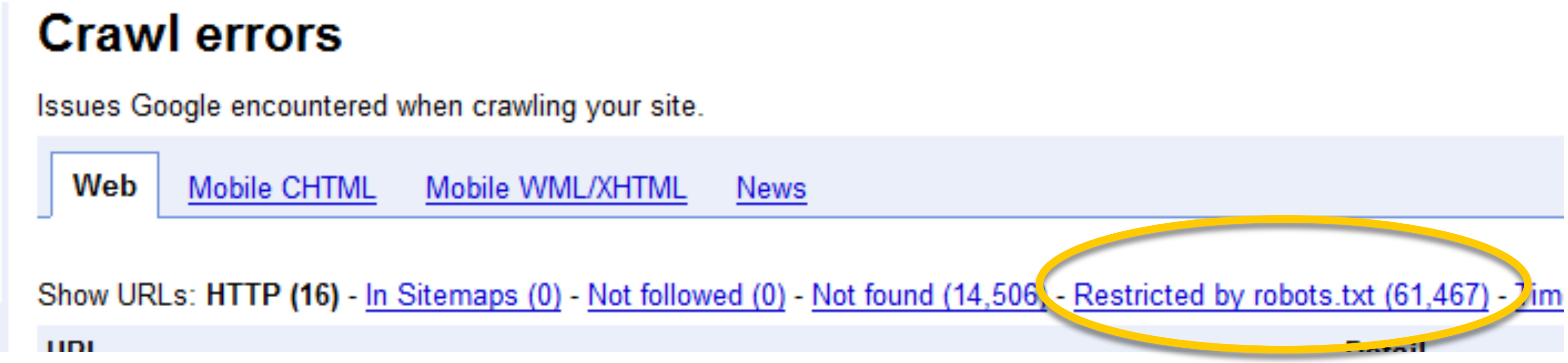
Issues Google encountered when crawling your site.

Show URLs: **HTTP (16)** - In Sitemaps (0) - Not followed (0) - Not found (14,506) - Restricted by robots.txt (61,467) - Timed out (0) - Unreachable (981)

URL	Detail	Detected
http://content.lib.utah.edu/EHSL-FBWNOC	4xx error	May 17, 2010
http://content.lib.utah.edu/EHSL-FBWNOC/	403 error	May 17, 2010

- □ Page Forbidden (401 errors)
- □ User Not Authorized (403 errors)
- □ Network Unreachable (5xx errors)
- □ Page Not Found (404 errors)

Eliminate sitemap & robots.txt conflicts



Crawl errors

Issues Google encountered when crawling your site.

Show URLs: HTTP (16) - In Sitemaps (0) - Not followed (0) - Not found (14,506) - Restricted by robots.txt (61,467) - Tim

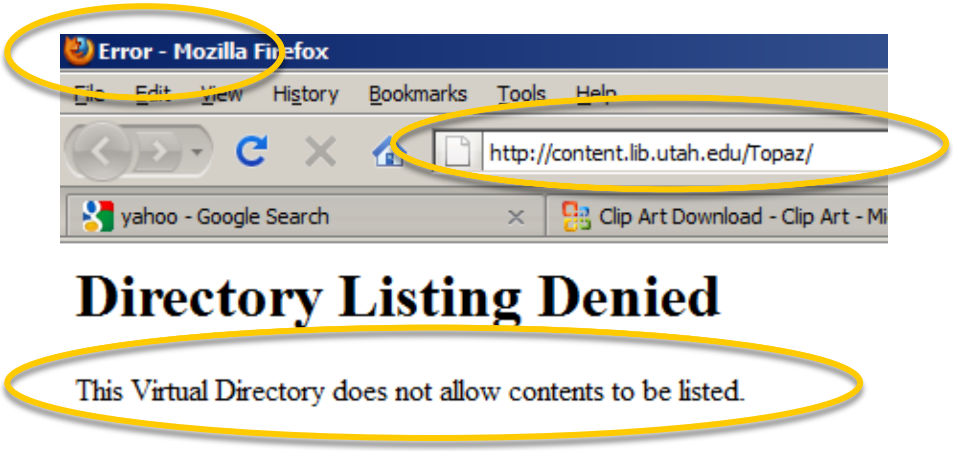
Robots.txt

```
User-agent: *
Disallow: /dmscripts/
Disallow: /cdm4/admin/
Disallow: /cdm4/client/
Disallow: /cdm4/cqr/
Disallow: /cdm4/images/
Disallow: /cdm4/includes/
Disallow: /cdm4/jscripts/
Disallow: /cdm-diagnostics/
Disallow: /cgi-bin/
Disallow: /images/
Disallow: /u/
```

Sitemap

http://content.lib.utah.edu/**cgi-bin/**
browseresults.exe?CISOROOT=/DC_Beckwith

Address errors and don't leave their customers stranded!


Error - Mozilla Firefox
File Edit View History Bookmarks Tools Help
http://content.lib.utah.edu/Topaz/
yahoo - Google Search
Clip Art Download - Clip Art - Mi

Directory Listing Denied

This Virtual Directory does not allow contents to be listed.

Low Trust Example

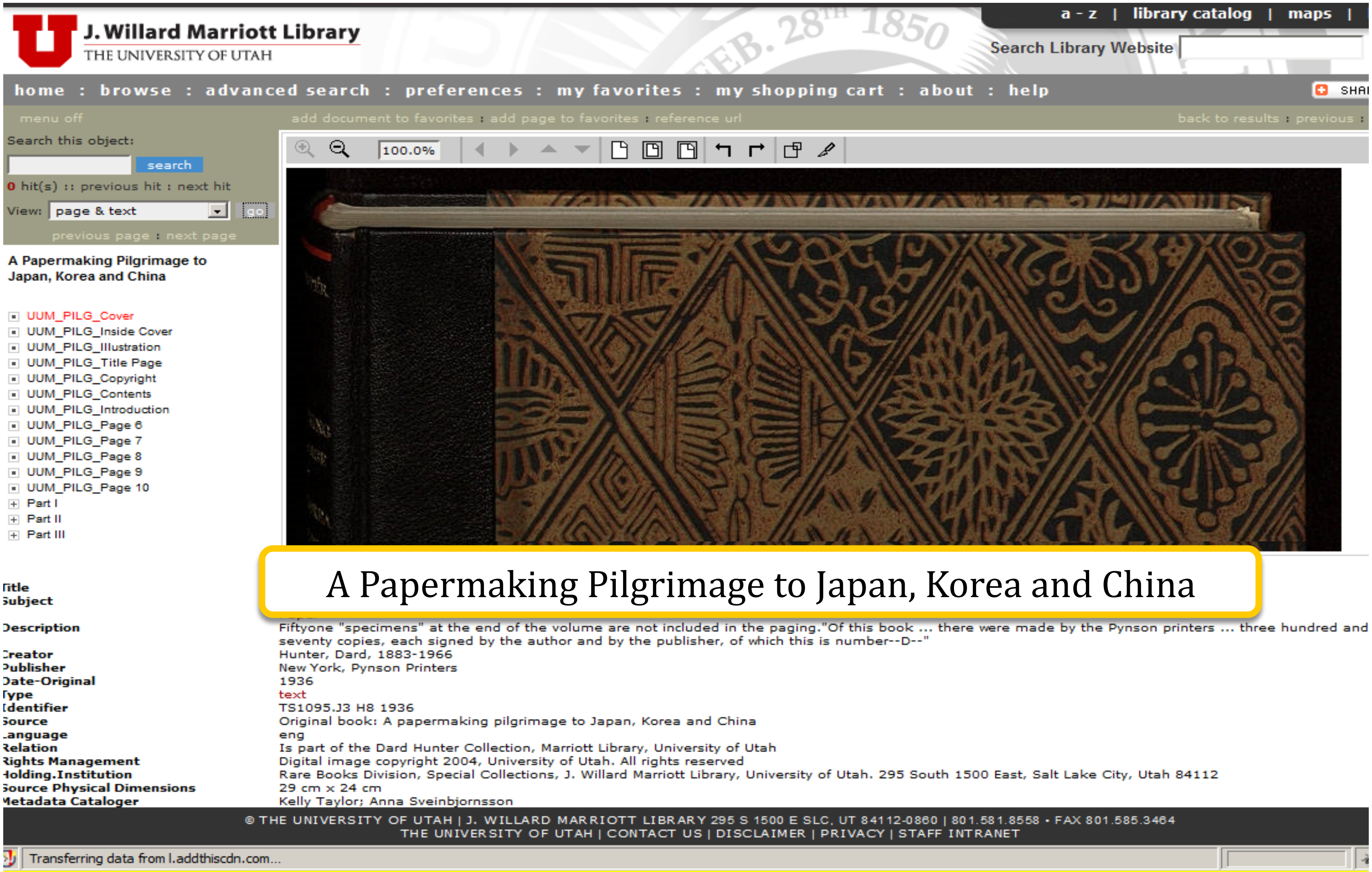
403 Error

How to Fix It	Example
Inform the Client Browser	<title>HTTP 403 Error</title> <meta HTTP-EQUIV = "Refresh" CONTENT = "8; URL =/"> <meta NAME="robots" CONTENT="NOINDEX,NOFOLLOW">

Inform the Search Engine	<?php
	header("HTTP/1.1 403 Forbidden);
	header("Location: http:// content.lib.utah.edu/");
	?>

Inform Their Customer	<p>The page you requested is no longer available or has been moved. </p> <p>You will be taken to our opening home page within the next 5 seconds. </p>

Provide path with context using simple URLs


J. Willard Marriott Library
THE UNIVERSITY OF UTAH
a - z | library catalog | maps |
Search Library Website
home : browse : advanced search : preferences : my favorites : my shopping cart : about : help
SHA
menu off
add document to favorites : add page to favorites : reference url
back to results : previous :
Search this object:
search
0 hit(s) :: previous hit : next hit
View: page & text
go
previous page : next page
100.0%

A Papermaking Pilgrimage to Japan, Korea and China

- UUM_PILG_Cover
- UUM_PILG_Inside Cover
- UUM_PILG_Illustration
- UUM_PILG_Title Page
- UUM_PILG_Copyright
- UUM_PILG_Contents
- UUM_PILG_Introduction
- UUM_PILG_Page 6
- UUM_PILG_Page 7
- UUM_PILG_Page 8
- UUM_PILG_Page 9
- UUM_PILG_Page 10
- Part I
- Part II
- Part III

A Papermaking Pilgrimage to Japan, Korea and China

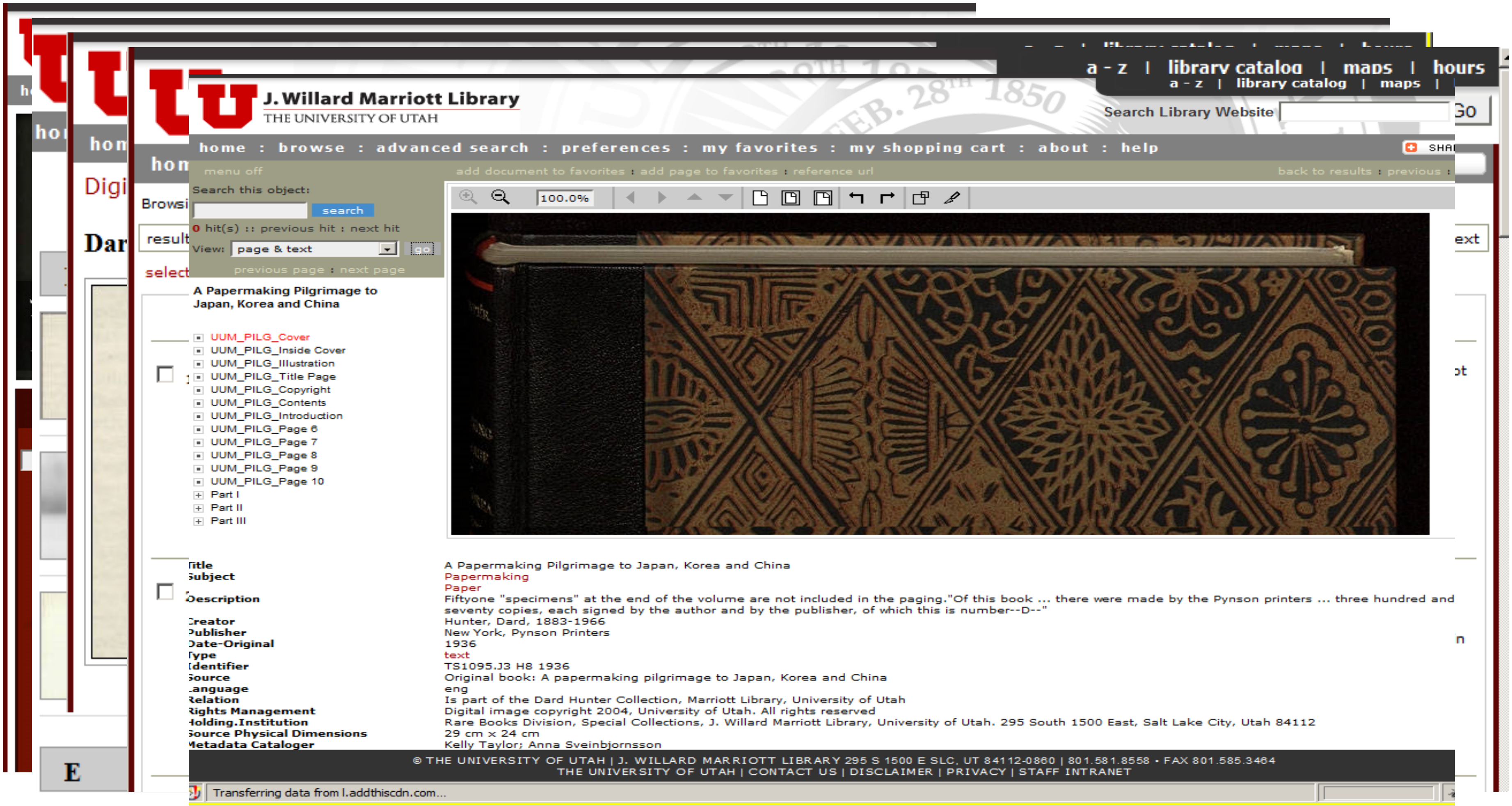
ʼitle	
ʼubject	
Description	Fiftyone "specimens" at the end of the volume are not included in the paging."Of this book ... there were made by the Pynson printers ... three hundred and seventy copies, each signed by the author and by the publisher, of which this is number--D--"
ʼreator	Hunter, Dard, 1883-1966
ʼublisher	New York, Pynson Printers
ʼate-Original	1936
ʼype	text
ʼdentifier	TS1095.J3 H8 1936
ʼource	Original book: A papermaking pilgrimage to Japan, Korea and China
ʼanguage	eng
ʼelation	Is part of the Dard Hunter Collection, Marriott Library, University of Utah
ʼights Management	Digital image copyright 2004, University of Utah. All rights reserved
ʼolding.Institution	Rare Books Division, Special Collections, J. Willard Marriott Library, University of Utah. 295 South 1500 East, Salt Lake City, Utah 84112
ʼource Physical Dimensions	29 cm x 24 cm
ʼetadata Cataloger	Kelly Taylor; Anna Sveinbjornsson

© THE UNIVERSITY OF UTAH | J. WILLARD MARRIOTT LIBRARY 295 S 1500 E SLC, UT 84112-0860 | 801.581.8558 • FAX 801.585.3464
THE UNIVERSITY OF UTAH | CONTACT US | DISCLAIMER | PRIVACY | STAFF INTRANET

Transferring data from l.addthiscdn.com...

Provide path with context using simple URLs

http://content.lib.utah.edu/cdm4/document.php?CISOROOT=/DardHunter&CISOPTR=1919


a - z | library catalog | maps | hours
J. Willard Marriott Library
THE UNIVERSITY OF UTAH
Search Library Website
Go
home : browse : advanced search : preferences : my favorites : my shopping cart : about : help
menu off
add document to favorites : add page to favorites : reference url
back to results : previous :
Search this object:
search
0 hit(s) :: previous hit : next hit
View: page & text
go
previous page : next page
100.0%
Digi
Browsi
result
select
Dar
E
ext
ot
n

A Papermaking Pilgrimage to Japan, Korea and China

- UUM_PILG_Cover
- UUM_PILG_Inside Cover
- UUM_PILG_Illustration
- UUM_PILG_Title Page
- UUM_PILG_Copyright
- UUM_PILG_Contents
- UUM_PILG_Introduction
- UUM_PILG_Page 6
- UUM_PILG_Page 7
- UUM_PILG_Page 8
- UUM_PILG_Page 9
- UUM_PILG_Page 10
- Part I
- Part II
- Part III

Title	A Papermaking Pilgrimage to Japan, Korea and China
Subject	Papermaking Paper
Description	Fiftyone "specimens" at the end of the volume are not included in the paging."Of this book ... there were made by the Pynson printers ... three hundred and seventy copies, each signed by the author and by the publisher, of which this is number--D--"
Creator	Hunter, Dard, 1883-1966
Publisher	New York, Pynson Printers
Date-Original	1936
Type	text
Identifier	TS1095.J3 H8 1936
Source	Original book: A papermaking pilgrimage to Japan, Korea and China
Language	eng
Relation	Is part of the Dard Hunter Collection, Marriott Library, University of Utah
Rights Management	Digital image copyright 2004, University of Utah. All rights reserved
Holding.Institution	Rare Books Division, Special Collections, J. Willard Marriott Library, University of Utah. 295 South 1500 East, Salt Lake City, Utah 84112
Source Physical Dimensions	29 cm x 24 cm
Metadata Cataloger	Kelly Taylor; Anna Sveinbjornsson

© THE UNIVERSITY OF UTAH | J. WILLARD MARRIOTT LIBRARY 295 S 1500 E SLC, UT 84112-0860 | 801.581.8558 • FAX 801.585.3464
THE UNIVERSITY OF UTAH | CONTACT US | DISCLAIMER | PRIVACY | STAFF INTRANET

Transferring data from l.addthiscdn.com...

Agenda

- □ Assessment
- □ Phase 1: Start feedback loop
- □ Phase 2: Get indexed
- □ Phase 3: Increase visibility (in progress)
- □ Wrap-up

Multiple Dynamic URLs pointing to a single URI

- Example: same content had 2+ URLs
  - http://content.lib.utah.edu/u?/ir-main,5239
  - content.lib.utah.edu/cdm4/document.php?CISOROOT=/ir-main&CISOPTR=370&CISOSHOW=5239
- Implemented Canonical Link Element to clarify 500+ URL Parameters

Google Scholar Bibliographic Metadata

"Use Dublin Core tags (e.g., DC.title) as a last resort - they work poorly for journal papers...

- *Google Scholar Inclusion Guidelines for Webmasters*

Embed bibliographic metadata in HTML & full text PDF files

- Mapped Dublin Core to a Google supported HTML meta tag
  - Highwire Press (e.g., citation_title)
- Extended Dublin Core fields
  - Journal Title
  - Journal Volume
  - Journal Issue
  - Starting Page Number
  - Ending Page Number
- Link directly to existing Full Text PDF

Link data to establish context and improve visibility

- Apply Taxonomy Schemas
  - Glossary
  - Acronyms
- External Linking
  - Authors
  - Organizations
  - External Feeds
- Target Audience Segments with Declared Ontology‘s

Agenda

- □ Assessment
- □ Phase 1: Start feedback loop
- □ Phase 2: Get indexed
- □ Phase 3: Increase visibility (future)
- □ Wrap-up

Lessons Learned

- Search engines want to send users to content that solves users' problem, not just to metadata
- Establish trust
- Linking strategies enormously important
    - Chicken and egg problem
- Ensure metadata is unique and descriptive
    - Dublin Core too ambiguous
    - Different audiences use different vocabularies
- Accessibility standards good for SEO

Managing expectations

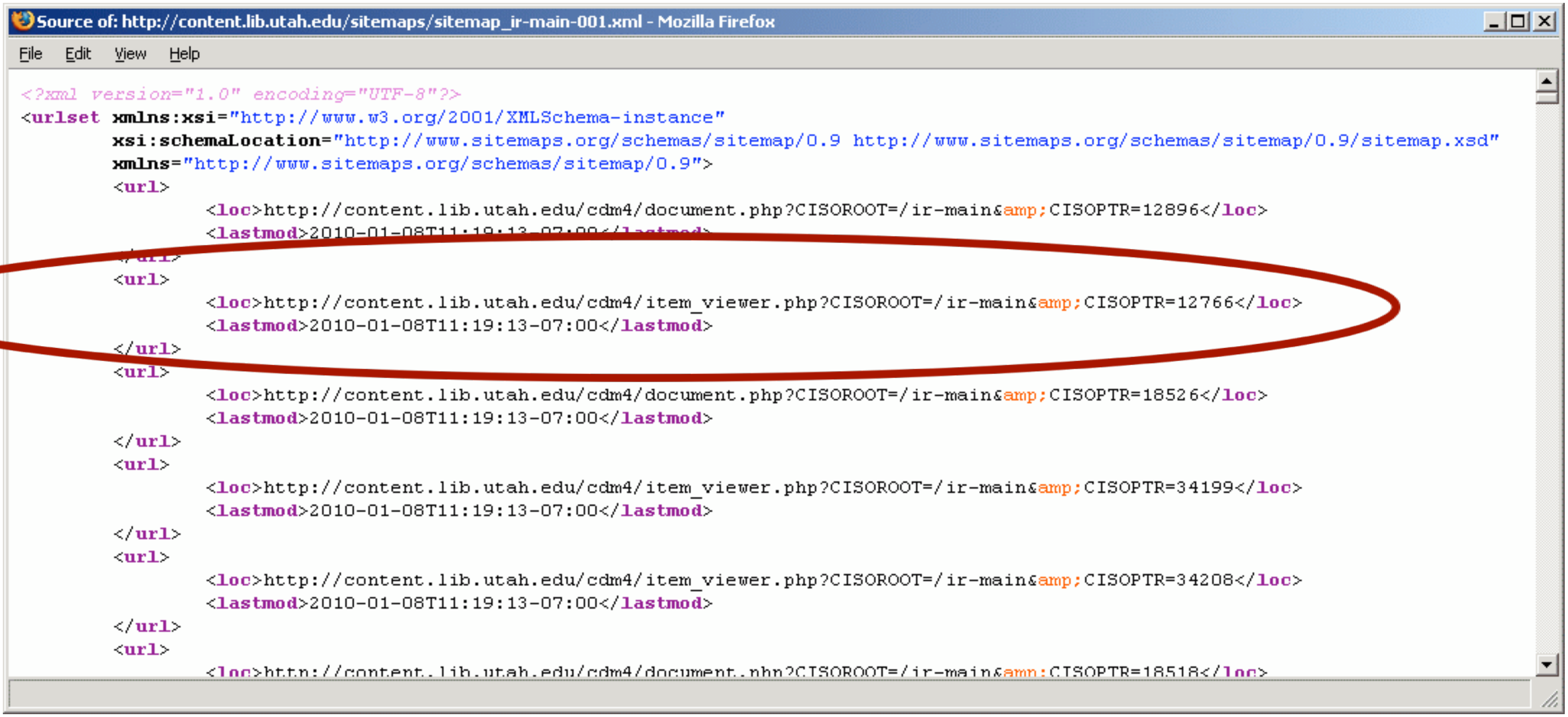
- SEO-SEM is a long-term strategy that requires constant monitoring
- Build a good site that is useful to people and engines will find it
- Search engine is the customer
- Influence vendors to add SEO features into products

Q&A

- Kenning Arlitsch
  - Associate Director for IT Services, Univ of Utah
  - kenning.arlitsch@utah.edu
- Sandra McIntyre
  - Program Director, Mountain West Digital Library
  - sandra.mcintyre@utah.edu
- Patrick O'Brien
  - Principal, RevX Corporation
  - patrick@revxcorp.com

Google Sitemap – example

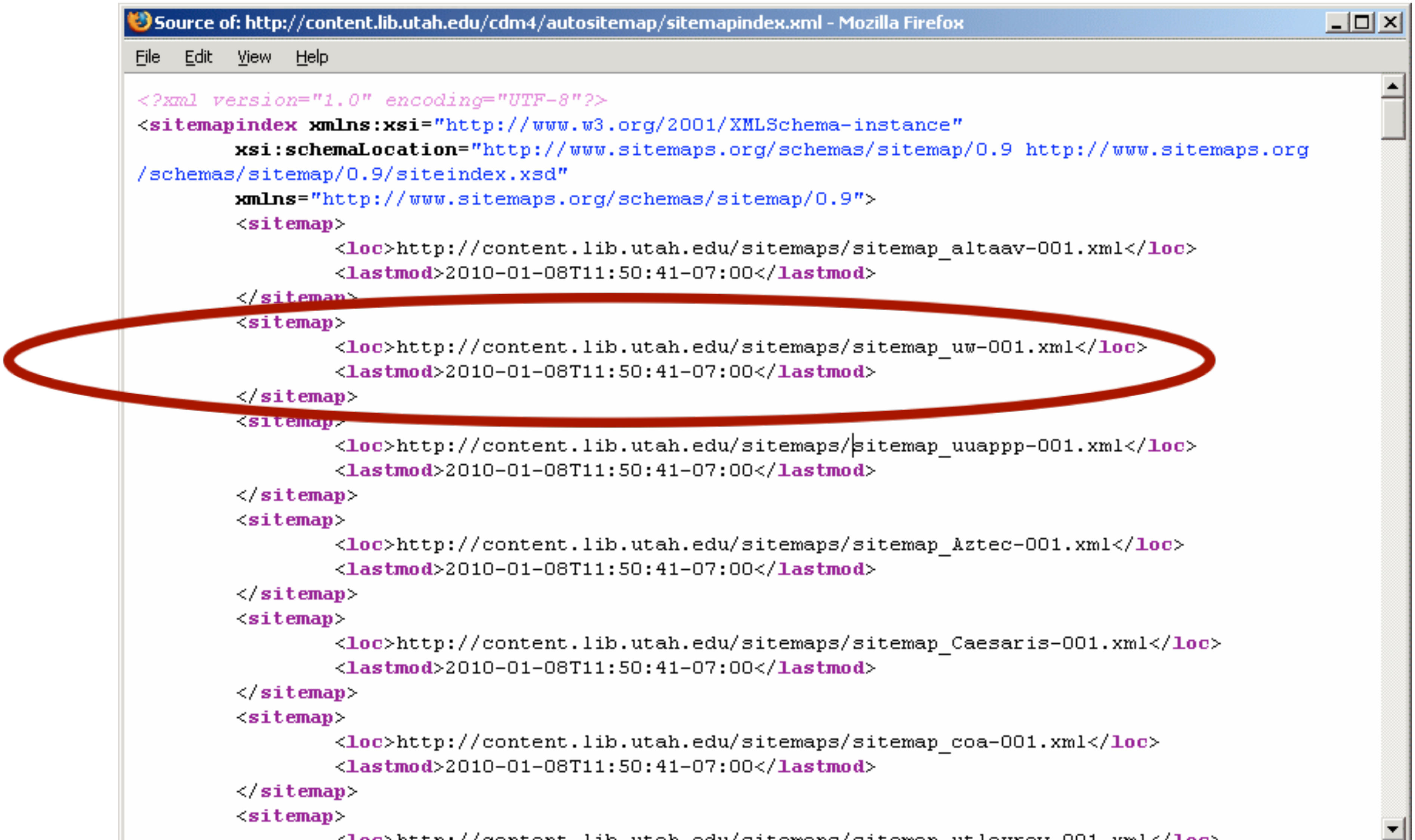
http://content.lib.utah.edu/sitemaps/sitemap_ir-main-001.xml


Source of: http://content.lib.utah.edu/sitemaps/sitemap_ir-main-001.xml - Mozilla Firefox
File
Edit
View
Help

```
<?xml version="1.0" encoding="UTF-8"?>
<urlset xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
        xsi:schemaLocation="http://www.sitemaps.org/schemas/sitemap/0.9 http://www.sitemaps.org/schemas/sitemap/0.9/sitemap.xsd"
        xmlns="http://www.sitemaps.org/schemas/sitemap/0.9">
        <url>
                <loc>http://content.lib.utah.edu/cdm4/document.php?CISOROOT=/ir-main&amp;CISOPTR=12896</loc>
                <lastmod>2010-01-08T11:19:13-07:00</lastmod>
        </url>
        <url>
                <loc>http://content.lib.utah.edu/cdm4/item_viewer.php?CISOROOT=/ir-main&amp;CISOPTR=12766</loc>
                <lastmod>2010-01-08T11:19:13-07:00</lastmod>
        </url>
        <url>
                <loc>http://content.lib.utah.edu/cdm4/document.php?CISOROOT=/ir-main&amp;CISOPTR=18526</loc>
                <lastmod>2010-01-08T11:19:13-07:00</lastmod>
        </url>
        <url>
                <loc>http://content.lib.utah.edu/cdm4/item_viewer.php?CISOROOT=/ir-main&amp;CISOPTR=34199</loc>
                <lastmod>2010-01-08T11:19:13-07:00</lastmod>
        </url>
        <url>
                <loc>http://content.lib.utah.edu/cdm4/item_viewer.php?CISOROOT=/ir-main&amp;CISOPTR=34208</loc>
                <lastmod>2010-01-08T11:19:13-07:00</lastmod>
        </url>
        <url>
                <loc>http://content.lib.utah.edu/cdm4/document.php?CISOROOT=/ir-main&amp;CISOPTR=18518</loc>
```

Sitemap Index - example

http://content.lib.utah.edu/cdm4/autositemap/sitemapindex.xml


Source of: http://content.lib.utah.edu/cdm4/autositemap/sitemapindex.xml - Mozilla Firefox
File
Edit
View
Help

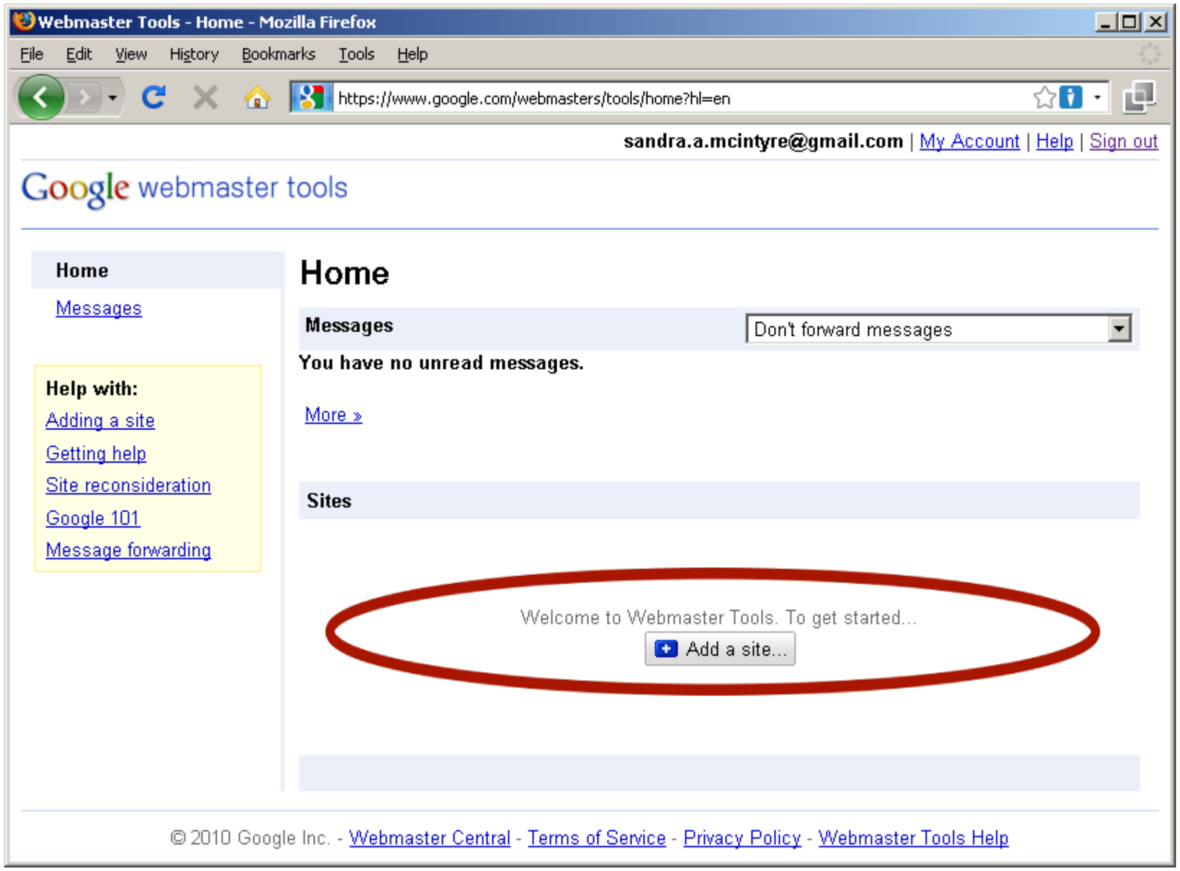
```
<?xml version="1.0" encoding="UTF-8"?>
<sitemapindex xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
        xsi:schemaLocation="http://www.sitemaps.org/schemas/sitemap/0.9 http://www.sitemaps.org
/schemas/sitemap/0.9/siteindex.xsd"
        xmlns="http://www.sitemaps.org/schemas/sitemap/0.9">
        <sitemap>
                <loc>http://content.lib.utah.edu/sitemaps/sitemap_altaav-001.xml</loc>
                <lastmod>2010-01-08T11:50:41-07:00</lastmod>
        </sitemap>
        <sitemap>
                <loc>http://content.lib.utah.edu/sitemaps/sitemap_uw-001.xml</loc>
                <lastmod>2010-01-08T11:50:41-07:00</lastmod>
        </sitemap>
        <sitemap>
                <loc>http://content.lib.utah.edu/sitemaps/sitemap_uuappp-001.xml</loc>
                <lastmod>2010-01-08T11:50:41-07:00</lastmod>
        </sitemap>
        <sitemap>
                <loc>http://content.lib.utah.edu/sitemaps/sitemap_Aztec-001.xml</loc>
                <lastmod>2010-01-08T11:50:41-07:00</lastmod>
        </sitemap>
        <sitemap>
                <loc>http://content.lib.utah.edu/sitemaps/sitemap_Caesaris-001.xml</loc>
                <lastmod>2010-01-08T11:50:41-07:00</lastmod>
        </sitemap>
        <sitemap>
                <loc>http://content.lib.utah.edu/sitemaps/sitemap_coa-001.xml</loc>
                <lastmod>2010-01-08T11:50:41-07:00</lastmod>
        </sitemap>
        <sitemap>
                <loc>http://content.lib.utah.edu/sitemaps/sitemap_utlawrev-001.xml</loc>
```

Step 1: Create Sitemaps and Index

- According to the protocol at http://www.sitemaps.org:
  - Create a Sitemap file for each collection.
  - Create a Sitemap Index file.

Step 2: Webmaster Tools Registration

- Register (free) with Google Webmaster Tools at http://www.google.com/webmasters/tools

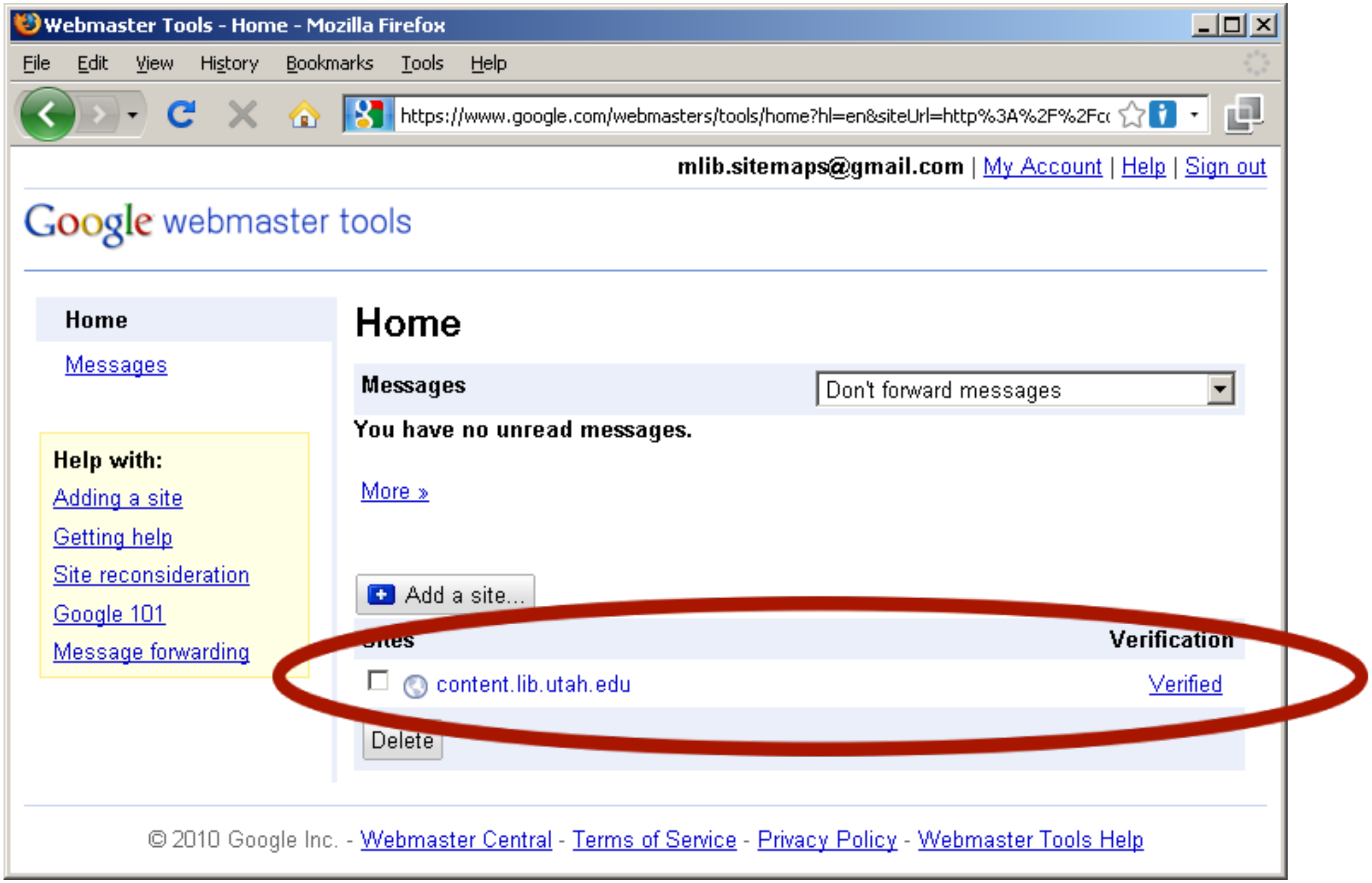

Webmaster Tools - Home - Mozilla Firefox
File Edit View History Bookmarks Tools Help
https://www.google.com/webmasters/tools/home?hl=en
sandra.a.mcintyre@gmail.com | My Account | Help | Sign out
Google webmaster tools
Home
Messages
Help with:
Adding a site
Getting help
Site reconsideration
Google 101
Message forwarding
Home
Messages
Don't forward messages
More »
Sites
Add a site...

You have no unread messages.

Welcome to Webmaster Tools. To get started...

© 2010 Google Inc. - Webmaster Central - Terms of Service - Privacy Policy - Webmaster Tools Help

Step 2: Webmaster Tools Registration



Webmaster Tools - Home - Mozilla Firefox
File Edit View History Bookmarks Tools Help
https://www.google.com/webmasters/tools/home?hl=en&siteUrl=http%3A%2F%2Fc
mlib.sitemaps@gmail.com | My Account | Help | Sign out
Google webmaster tools
Home
Messages
Help with:
Adding a site
Getting help
Site reconsideration
Google 101
Message forwarding

Home

Messages Don't forward messages

You have no unread messages.

More »

Add a site...

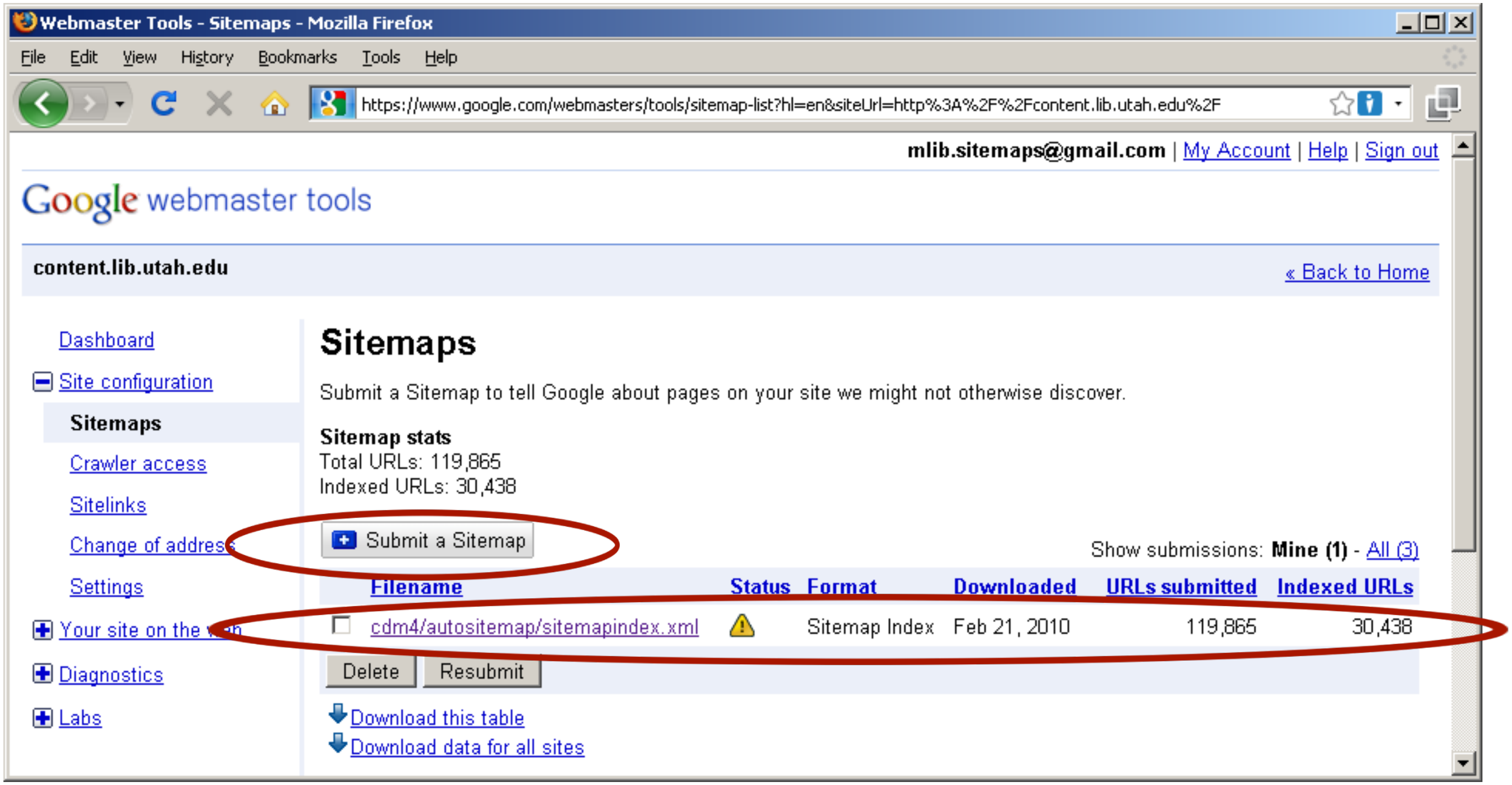
Sites	Verification
□ content.lib.utah.edu	Verified

Delete

© 2010 Google Inc. - Webmaster Central - Terms of Service - Privacy Policy - Webmaster Tools Help

Step 3: Inform Google

- Step 3A: Submit the address of Sitemap Index file on Webmaster Tools.



Webmaster Tools - Sitemaps - Mozilla Firefox
File Edit View History Bookmarks Tools Help
https://www.google.com/webmasters/tools/sitemap-list?hl=en&siteUrl=http%3A%2F%2Fcontent.lib.utah.edu%2F
mlib.sitemaps@gmail.com | My Account | Help | Sign out
Google webmaster tools
content.lib.utah.edu
« Back to Home
Dashboard
Site configuration
Sitemaps
Crawler access
Sitelinks
Change of addres
Settings
Your site on the w
Diagnostics
Labs

Sitemaps

Submit a Sitemap to tell Google about pages on your site we might not otherwise discover.

Sitemap stats
Total URLs: 119,865
Indexed URLs: 30,438

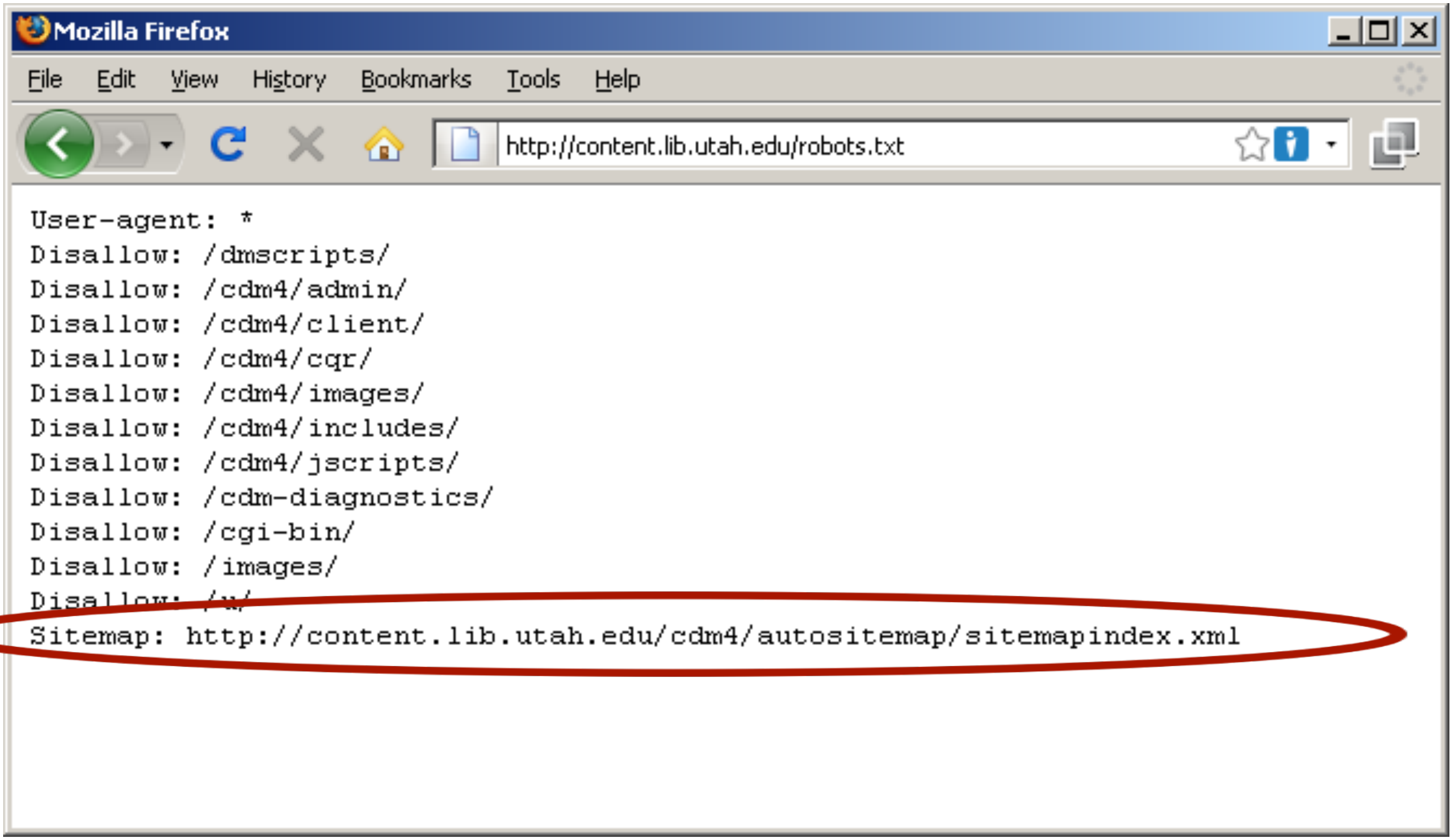
Submit a Sitemap
Show submissions: Mine (1) - All (3)

	Filename	Status	Format	Downloaded	URLs submitted	Indexed URLs
□	cdm4/autositemap/sitemapindex.xml	⚠	Sitemap Index	Feb 21, 2010	119,865	30,438

Delete
Resubmit
Download this table
Download data for all sites

Step 3: Inform Google

- □ Step 3B: Modify the robots.txt file at the root of your CONTENTdm server to specify the location of the Sitemaps Index.



```
User-agent: *
Disallow: /dmscripts/
Disallow: /cdm4/admin/
Disallow: /cdm4/client/
Disallow: /cdm4/cqr/
Disallow: /cdm4/images/
Disallow: /cdm4/includes/
Disallow: /cdm4/jscripts/
Disallow: /cdm-diagnostics/
Disallow: /cgi-bin/
Disallow: /images/
Disallow: /u/
Sitemap: http://content.lib.utah.edu/cdm4/autositemap/sitemapindex.xml
```