

# Finding Hidden Gems in Online Databases & Repositories

Emerging Technology for Scholars Series

Session 4

Presented by Bill Warters

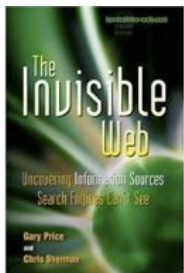
(OTL Faculty Fellow)

See Companion Site (visit as guest)

<http://snipurl.com/emergingtech>

1

## Accessing the "Invisible" Web



- ④ The "visible web" is what you see in the results pages from general web search engines. It's also what you see in almost all subject directories.
- ④ The "invisible web" or "deep web" is what you cannot retrieve ("see") in the search results and other links contained in these types of tools.

2

# Search Engine Improvements

Certain kinds of pages used to be invisible but can now be found in most search engine results:

- Pages in non-HTML formats (pdf, Word, Excel, Corel suite, etc.) are "translated" into HTML now in most search engines and can "seen" in search results.
- Script-based pages, whose links contain a "?" or other script coding, no longer cause most search engines to exclude them.
- Pages generated dynamically by other types of database software (e.g., Active Server Pages, Cold Fusion) can be indexed if there is a stable URL somewhere that search engine spiders can find. Once these were largely shunned by search engines.

3

## Search engines still cannot type or think...



- If access to a web pages requires typing, web crawlers encounter a barrier they cannot go beyond. They cannot search online catalogs and they cannot enter a password or login.
- The "Deep Web" thus is now largely that which is **hidden inside searchable databases**
- Thousands of public-record, official, & special-purpose databases containing government, financial, logistical, & other types of info that is needed to answer very specific inquiries

4

# Using the Search Engines to Find "hidden" content

- ④ Use Google's advanced search features to specify specialized file type (e.g. powerpoint) and narrow your domain type (e.g. .edu)
  - "workplace violence" site:.edu filetype:ppt
- ④ Use search engines to locate useful databases by searching a **subject term** and the word **"database"**.
  - toxic chemicals database
  - plane crash database

5

## Some Nice Examples...

<http://chnm.gmu.edu/tools/syllabi/>

The screenshot shows the website for the Center for History & New Media's Syllabus Finder. The page features a search bar at the top right with the text "Search: [input] [GO]". Below the search bar is a navigation menu with links for "home", "projects", "tools", "resources", and "news". On the left side, there is a "Tools" menu with links for "Zotero", "Syllabus Finder", "Web Scrapbook", "Survey Builder", "Timeline Builder", "Poll Builder", "Scribe", "H-Bar", "Tools Center", "Syllabus Finder", "Search Syllabi", "About this Tool", "Help", and "Tips & FAQ". The main content area features a large image of a person with their arms raised, with the text "Syllabus Finder" written below it. Below the image is a search input field with the placeholder text "Enter keywords, phrases, names, or IDs:" and a "Search" button. Below the search field, the text reads "Searching 776,196 syllabi at the Center for History and New Media and over 500,000 syllabi via Google." and "From Dan Cohen's blog: the most popular syllabi in history and philosophy." At the bottom of the page, it says "Powered by Center for History and New Media databases and the Google Web API."

Center for History & New Media  
Syllabus Finder

6

# Google Video Search using genre:educational

Publishers have the option to select 38 different genres they'd like to categorize their video in. Many of these categories are not linked anywhere on the google video website.

- |   |   |
|---|---|
| <ul style="list-style-type: none"><li>◦ Educational</li><li>◦ Instructional</li><li>◦ Stock Footage</li><li>◦ Comedy</li><li>◦ Mature Adult</li><li>◦ Science and Technology</li><li>◦ Ads and Promotions</li><li>◦ News</li><li>◦ Animation</li><li>◦ Sports</li><li>◦ Home Video</li><li>◦ Documentary</li><li>◦ Business</li><li>◦ Dance</li><li>◦ Drama</li><li>◦ Foreign</li><li>◦ Gaming</li><li>◦ Horror</li><li>◦ Movie trailer</li></ul> | <ul style="list-style-type: none"><li>◦ Nature</li><li>◦ Political</li><li>◦ Religious</li><li>◦ Romance</li><li>◦ Special Interest</li><li>◦ Thriller</li><li>◦ Travel</li><li>◦ TV Show</li><li>◦ Western</li><li>◦ Movie Short</li><li>◦ Movie Feature</li><li>◦ Faith and Spirituality</li><li>◦ Gay and Lesbian</li><li>◦ Science Fiction &amp; Fantasy</li><li>◦ Action and Adventure</li><li>◦ Children and Family</li><li>◦ Art and Experimental</li><li>◦ Health And Fitness</li><li>◦ Independent</li></ul> |
|---|---|

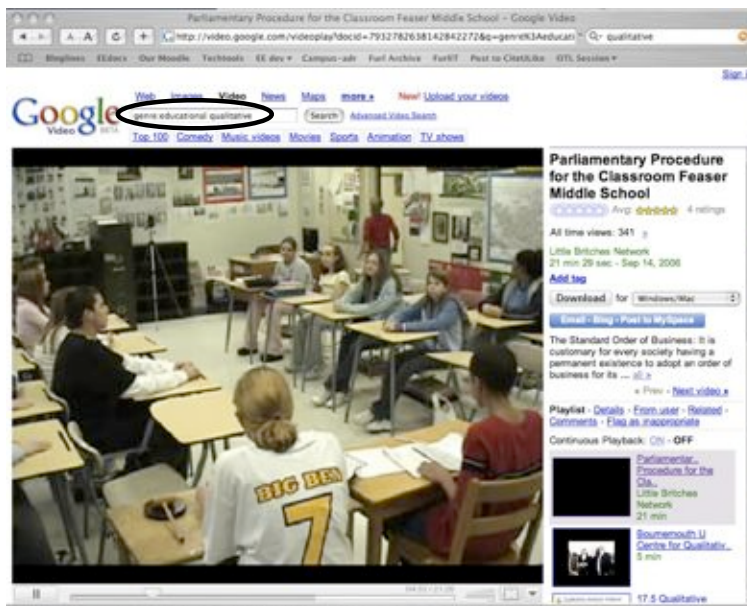
1 - Go to Google Video - <http://video.google.com>

2 - In the search box, type: "genre:educational"

3 - Follow it with anything you want to find video for (e.g. "lecture" or "chemistry")

7

# Google Video Search Results Example



8

# <http://search.singingfish.com/> Singing Fish Audio & Video Search



9

## Use Reliable Curated Directories of Content

- ① Librarians Index of the Internet

<http://lii.org/>



- ① Infomine

<http://infomine.ucr.edu/>



- ① AcademicInfo

<http://www.academicinfo.net/>

- ① INTUTE

<http://www.intute.ac.uk/>



10

# Open Archives Repositories & Institutional Collections

- There has been a broad movement toward Open Access of Scholarly Publications due to high costs of journal subscriptions and desire for higher profile for faculty and staff expertise.
- Wayne State University has a growing digital archives collection and an institutional repository that is just waiting for your content  
[http://www.lib.wayne.edu/geninfo/about/scholarly\\_communications/](http://www.lib.wayne.edu/geninfo/about/scholarly_communications/)
- Archives around the world have adopted a shared system for tagging their content - a Protocol for Metadata Harvesting (OAI-PMH)

11

## OAster



<http://oaister.umdl.umich.edu/o/oaister/>

12



# So...What Are Learning Objects?

"The main idea of 'learning objects' is to break educational content down into small chunks that can be reused in various learning environments, in the spirit of object-oriented programming"

-- David A. Wiley



## Learning Object Photographed

Researchers were surprised to find the first photograph of the so-named "learning object." Previously thought to be a strictly a theoretical construct, the Learning Object was spotted on the campus of The Ohio State University (pictured behind the bush).

"We didn't think we would actually find a Learning Object," says scientist J.R. Morgan, "but we found that by supplying it with some Metadata, they would easily come to us. Once we understood that, we were able to find Learning Objects everywhere." In fact, the scientists also discovered previously hidden Learning Objects that were only revealed through careful metadata placement (see related story: "Was OSU's National Title Win Due to Learning Objects?")

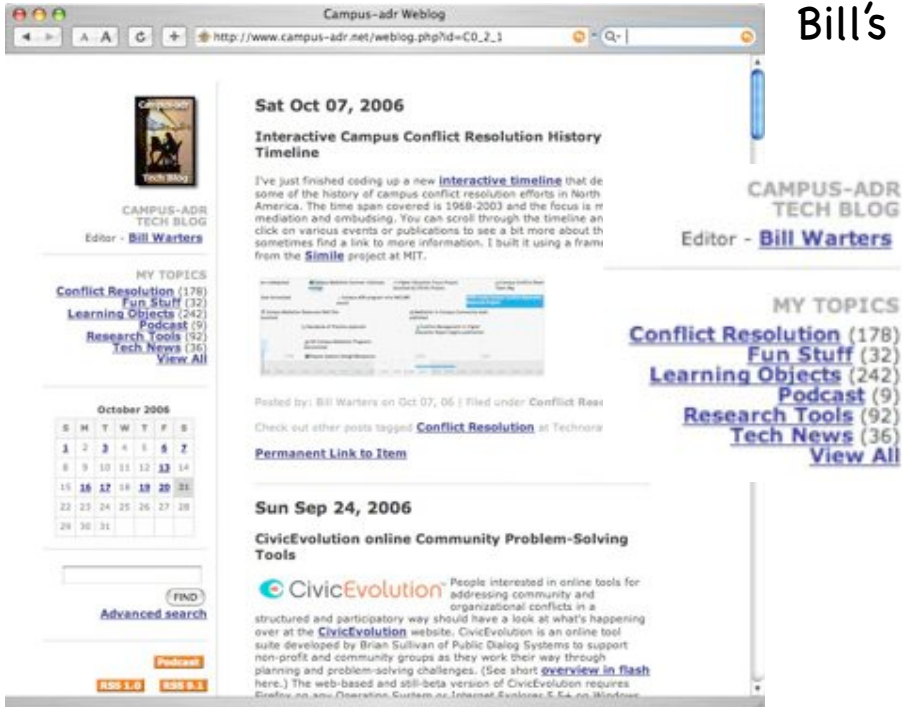
13

The Idea of a Learning Objects can be complex...



14

Or very simple and imprecise... For examples see Bill's blog:



<http://www.campus-adr.net/weblog.php>

15

## LO Examples from Wisc-Online Resource Center

<http://www.wisc-online.com/>



16



# Great Listing of Learning Objects Collections

See Handout...

[http://www.uwm.edu/Dept/CIE/AOP/LO\\_collections.html](http://www.uwm.edu/Dept/CIE/AOP/LO_collections.html)

17

## A Few Favorites...

- Blue Web'n

<http://www.kn.sbc.com/wired/bluwebn/>



- Merlot

<http://www.merlot.org/merlot/index.htm>



- Canadian Broadcasting Corp Archives

<http://archives.cbc.ca/>



- E-Clips from Cornell University

<http://eclips.cornell.edu/>



18

# Keeping Track of your Good Stuff...

- Remember you can use Social Bookmarking Tools that we explored in an earlier session
  - FURL
  - Delicious
  - Blog about them...