Search Multiple Data Websites Simultaneously Using APIs

Jeff Essic
Data Services Librarian
North Carolina State Univ. Libraries
Raleigh, NC

2016 Esri Education GIS Conference
GIS at NC State University

ArcGIS installed in all campus labs

UNC system-wide site license for ArcGIS personal computer install

Virtual Campus Course Enrollments, 7/2015 - 6/2016: 2763

ArcGIS Online Members: 630
NCSU Center for Geospatial Analytics

42 GIS Faculty & 11 Affiliates, 157 Grad Students

500+ Enrollments in GIS courses per academic year (200 each semester in Intro. to GIST)

Masters GIST, Minors, Cert., PhD in development

Photo from http://gis.ncsu.edu
NCSU Libraries Geospatial Data Services

5 total GIS and Visualization workstations at Hill and Hunt Libraries

1 Data Services Librarian, within Reference & Information Services
Libraries’ GIS Data Services History

Mid 90’s - received a lot of local and state data on CDs

Late 90’s - “GIS Lookup” catalog developed using Access database and perl scripts

Mid 2000’s - Transitioned GIS Lookup to dynamic catalog with MySQL/php

Developed batch XML parsing tool to import XML metadata fields into MySQL

Most popular: Esri Data and Maps (delivered on DVDs) and NC OneMap datasets
GIS Lookup - getting behind

2007 - my son born. No more leisure time for cataloging. GIS Lookup started getting behind.

ESRI Data and Maps for ArcGIS, ArcGIS StreetMap USA (2006)

Overview | Networked Access | Library Access | Restrictions | Previous Versions

Please Note: The 2006 version of Esri Data and Maps was the last version that was extensively cataloged into the Libraries' GIS Lookup database. More recent data are available, but not cataloged. See Esri Data and Maps general webpage.

https://www.lib.ncsu.edu/gis/esri2006.html
Esri Data & Maps via ArcGIS Online

May 25, 2010

Esri Launches ArcGIS.com and ArcGIS Explorer Online

New Resources Give Users the Ability to Share GIS Content and Build Communities

Redlands, California—Esri announces the launch of ArcGIS.com, a Web site for finding and sharing geographic information system (GIS) content, organizing geographic information into portfolios, and building communities. ArcGIS Explorer is now available online.

Inspiration!

The Puzzle

Challenges:
• Only 1 GIS Librarian/cataloger (me)
• Late 2000’s budget cuts (“do more with less”)
• Time costs for linking and cataloging > Value

Numerous, widely distributed data silos
• data are already cataloged
• stable and free download access
• why duplicate work?
• how to point people in the right direction?

Needed: a custom data search aggregator
Solution: External Data Sources

results window

http://www.lib.ncsu.edu/gis/search/search.php?q=streets
## External Data Sources

Search for **streets** - The top results (relevance sort) are listed below:

### nc.ONE.map
- 14 Results beginning with...
  - Primary and Secondary Routes, 1st Quarter 2015 - NC Department of Transportation
  - Primary and Secondary Road Arcs, 1st Quarter 2015 - NC Department of Transportation
  - Road Characteristics, 1st Quarter 2015 - NC Department of Transportation

### ArcGIS Online
- 22141 Results beginning with...
  - Streets
  - World Street Map
  - Streets
  - GeoData.GOV

### Geo.Data.gov
- 2027 Results beginning with...
  - TIGER/Line Shapefile, 2014, Series Information for the All Roads County-based Shapefile
  - Great Chile Earthquake of May 22, 1960 - Anniversary Edition
  - TIGER Roads 2015

### Trimble
- 131 Results beginning with...
  - USGS DRG Topo Street, MS-LA
  - OpenStreetMap Planet
  - Neighborhoods: Asia

Links to search results on each site using query string
//BEGIN FUNCTION - searches the different data sites
function portal($type,$addr,$dirl,$site,$query,$focus,$url) {

    //Number of titles to display
    $numTitles = 3;

    $ch = curl_init();

    //READ THE FEED
    if ($type == "rest") {
        $chURL = "$addr$dirl/rest/distributed?rid=$site&start=1&max=$numTitles&orderBy=relevance&searchText=$query"
    } elseif ($type == "weo") {
        $chURL = "$addr/search.xml?per_page=$numTitles&use_facets=false&query=$query"
    } elseif ($type == "json") {
        $chURL = "$addr/api/3/action/package_search?q=$query&rows=$numTitles&facet=on&facet.field=metadata_type&facets=on&facets.field=geospatial"
    }
    //echo $chURL; //to debug

    curl_setopt($ch, CURLOPT_URL, $chURL);
    curl_setopt($ch, CURLOPT_HEADER, 0);
    curl_setopt($ch, CURLOPT_RETURNTRANSFER, 1);
    curl_setopt($ch, CURLOPT_CONNECTTIMEOUT, 300);
    curl_setopt($ch, CURLOPT_TIMEOUT, 300);
    $data = curl_exec($ch);
    curl_close($ch);

    //REPORT THE NUMBER
    if ($type == "rest") {
        $number = preg_replace("/\[D]/", ",", $data);
    } elseif ($type == "weo") {
        $number = strstr($data, 'total_pages', true);
        $number = strstr($number, 'total_entries=');
        $number = preg_replace("/\[D]/", ",", $number);
    } elseif ($type == "json") {
        $obj = json_decode($data, true);
        $number = $obj['result']['count'];
        $number = $obj['result']['facets']['metadata_type']['geospatial'];
    }
NC OneMap
$type = "rest";
$addr = "http://data.nconemap.gov.ua/";
$site = "local";
$dir1 = "geoportal";
$focus = "all";
echo '<tr><td align="center" style="padding-right: 10px;">'.'<img src="nconemap.gif" alt="NC OneMap" /></td><td><a href="'.$resultsLink.'">Search", "NC OneMap", "$query.");"><strong>portal($type,$addr,$dir1,$site,$query,$focus,$url);</strong>
echo '</td></tr>\n";

//ArcGIS.com
$type = "rest";
$addr = "http://gptogc.esri.com.ua/";
$dir1 = "geoportal";
$site = "arcgis.com";
$focus = "all";
$resultsLink = "http://www.arcgis.com/home/search.html?q=$query";
echo '<tr><td align="center" style="padding-right: 10px;">'.'<img src="arctgisonline.gif" alt="ArcGIS Online" /></td><td><a href="'.$resultsLink.'">ArcGIS.com", "$query.");"><strong>portal($type,$addr,$dir1,$site,$query,$focus,$url);</strong>
echo '</td></tr>\n";

//geo.data.gov
$type = "json";
$addr = "http://catalog.data.gov.ua/"
$dir1 = "";
$site = "";
$focus = "";
$url = "http://catalog.data.gov.ua/dataset?metadata_type=geospatial&metadata_type_limit=0&q=$query";
$resultsLink = "http://catalog.data.gov.ua/dataset?metadata_type=geospatial&metadata_type_limit=0&q=$query";
echo '<tr><td align="center" style="padding-right: 10px;">'.'<img src="geodatagov.gif" alt="geo.data.gov.ua" /></td><td><a href="'.$url.'">geo.data.gov.ua", "$query.");"><strong>portal($type,$addr,$dir1,$site,$query,$focus,$url);</strong>
echo '</td></tr>\n";

//Trimbledata.com
$type = "weo";
$addr = "http://market.trimbledata.com.ua/"
$dir1 = "";
<xml version="1.0" encoding="UTF-8">
    <channel>
      <title>Geoportal GeoRSS</title>
      <description>Most recently updated metadata documents.</description>
      <link>http://gptogc.esri.com/geoportal</link>
      <docs>http://www.rssboard.org/rss-specification</docs>
      <category>GeoRSS</category>
      <generator>Geoportal</generator>
      <copyright>Esri</copyright>
      <description>Open</description>
      <opensearch:totalResults>22146</opensearch:totalResults>
      <opensearch:startOffset>0</opensearch:startOffset>
      <opensearch:itemsPerPage>3</opensearch:itemsPerPage>
      <item>
        <title>World Street Map</title>
        <description><![CDATA[
This map was last updated June 2016. This worldwide street map presents highway-level data for the world down to 1:72k scale. Most of Canada, Japan, Europe, much of Russia; Austral...]]></description>
        <media:title><![CDATA[World Street Map]]></media:title>
        <media:content>http://gptogc.esri.com/geoportal/catalog/livedata/preview.page?uuid=3b93337983e94366db95e38a8629a autistic&url=http%3A%2F%2Fservices.arcgisonline.com%2Farcgis%2Frest%2Fservices%2FWorld_Street_Map%2FMaps%2Fdata%3F" rel="related" type="application/atom+xml" title="World Street Map"/>
        </item>
      </channel>
    </rss>
  </xml>
The parameter that controls which fields are returned in the solr query cannot be changed. CKAN always returns the matched datasets as dictionary objects.

"count": 2027,
"display_name": "geospatial",
"name": "geospatial",
"title": "TIGER/Line Shapefile, 2014, Series Information for the All Roads County-based Shapefile",
"type": "geospatial",
"format": "XML",
"tracking_summary": {
"total": 31,
"recent": 2,
"mime_type": "application/xml",
"package_count": 1,
"number_resources": 4,
"tags": []
}
This XML file does not appear to have any style information associated with it. The document tree is shown below.

```xml
<search current_page="1" per_page="3">  
  <facets type="array">  
    <total_entries="131" total_pages="44">  
      <item>  
        <price-type>FIXED</price-type>  
        <min-price type="decimal">0.0</min-price>  
        <max-price type="decimal">0.0</max-price>  
        <lowest-price type="decimal">0.0</lowest-price>  
        <highest-price type="decimal">0.0</highest-price>  
      </item>  
      <name>USGS DRG Topo Street, MS-LA</name>  
    </total_entries>  
  </facets>  
</search>
```

Example Topo detail and cultural features. The original, paper topo maps produced by the USGS have been scanned and georeferenced to create additional topos for this series and scale master: USGS DRG Topo 24k Master.

For more information, visit [USGS homepage](http://www.usgs.gov/).
Benefits

Expose multiple data collections / sources

95% cut in time spent cataloging

100% cut in stress over cataloging
Google Analytics

37,748 initial searches. Subsequent searches are tracked by the URL of the last search.

Something broke
Problems & Concerns

ArcGIS.com connection depends on a “sandbox” – gptogc.esri.com

• screen scraping may be a possibility if it goes away

Would be nice if ArcGIS.com individual dataset displays could provide more metadata (especially Dates)

“Relevance” sorting of search results are not reliable

Lack of error handling in the php code
Feedback and Suggestions

Are there any other similar services out there?

Ideas for improvement?

Contact:
Jeff Essic
jfessic@ncsu.edu
919-515-5698