

# **UP AND RUNNING WITH BENTO:**

**IMPLEMENTING BENTO SEARCH,  
DISCOVERY SERVICE AND HOMEPAGE  
UPDATE IN A SEMESTER**

**LAURA WIEGAND**

ASSOCIATE DIRECTOR, LIBRARY INFORMATION TECHNOLOGY  
AND DIGITAL STRATEGIES

**JASON FLEMING**

INFORMATION TECHNOLOGY LIBRARIAN

UNIVERSITY OF NORTH CAROLINA WILMINGTON

[HTTP://LIBRARY.UNCW.EDU](http://library.uncw.edu)

# TABLE OF CONTENTS

- Background
- Process
- Why a bento search?
- Technical solutions
- Impacts & future considerations

# QUESTION

*What would you like to get out of this presentation?*

**BACKGROUND**

# US

- Medium-sized academic library
- 13,914 Undergraduates, 1,826 Graduates
- 52 staff, 8 Library IT staff
- Systems:
  - ILS = Sierra
  - Discovery Service = Summon (was WCL)
  - Link Resolver = 360 Link (was III WAM)
  - CMS = Drupal

# YOU

- *Do you have a bento box search already?*
- *What systems (ILS, Discovery, etc.) do you use?*

**PROCESS**

# TIMELINE

- 2013 Began implementation of first discovery service: WorldCat Local
- August 2014 Fully launched WCL
- September 2014 Began review of discovery services
- November 2014 Concluded review & recommended to migrate to Summon
- December 2014 Began Summon implementation
- January 2015 Began Bento implementation
- April 2015 Beta versions released
- May 2015 went live with Bento search, 360 Link and Summon



# KEYS TO SPEEDY IMPLEMENTATION

- Project Manager
- Target dates
- Project tracking
- Larger advisory group & much smaller working groups
- Communication (heads up)
- Agreeing to iterative solutions
- Learn from and build on existing solutions

# DIVIDE AND CONQUER

- Coordinator of Discovery Services
- Task Force for Summon Implementation
  - Representatives from IT, Technical Services, Public Services
- Sub-groups
  - Summon Interface
  - 360 Link set-up
  - 360 Link interface
  - Catalog integration
  - Bento Search
  - Marketing/Outreach

# PROJECT TRACKING

| Summon Interface   | 360 Link Set-up   | 360 Link Interface  | Catalog   | Search box & bento   | Help/Marketing   |
|--|---|---|---|--|--|
| Due Date: 4/2/2015   | Due between 4/2-5/11  | Due between 4/2-5/11  | Export Due 5/18/2015  | Due Date: 4/2/2015   | Due Date: 5/11/2015  |
| Members: LKW, PF, AP, MER  | Members: LKW, JF, LS, GM, DP  | Members: LKW, MER, PF, JF, DP   | Members: LKW, GM, SB, JF, LS  | Members: LKW, PF, AP, GM, SB, JF   | Members: LKW, PF, AP, MER  |
| Interface configuration<br>Best bets<br>Database<br>Recommenders<br>Phrase translations<br>Facets<br>Default settings<br><a href="#">Summon</a>  | Setting holdings (journals, <a href="#">ebooks</a> )<br>Link resolver <a href="#">config</a><br>Migrate from <a href="#">webbridge</a> (due July)<br>Add print?<br><a href="#">Review checklist</a>                                 | Link resolver interface<br>Testing<br><a href="#">Example</a>   | Integration with Summon (metadata mapping, exporting)<br>Setting up daily exports (Due 3/1/2015)<br>Interface improvements (Due July)                               | Homepage search box<br>Bento results page (categories, display)<br><br><a href="#">Live page</a>   | Help page(s)<br>Tutorial(s) (Summon and Get it from UNCW?)<br>Marketing ( <a href="#">Proquest</a> will do a staff webinar)<br>Usability |
| <b>Status or next steps:</b><br>Done, except I need to add in Best Bets: Will add in <a href="#">dbs</a> that are not in database recommender set.<br>Note from meeting: Find a way to make it not so confusing to users about the database recommendations and best bet categories. | <b>Status or next steps:</b><br>Update from LS.<br><br>Need to test and review linking. (Gale)<br><br>If testing goes well, make plan to migrate from <a href="#">webbridge</a> for external linking.<br><br>Add print before fall. | <b>Status or next steps:</b><br>Done, except for to review interface for "no results found" page.<br><br>Usability testing. | <b>Status or next steps:</b><br>Records sent 3/9.<br><br>After import need to look at Location codes and MARC mapping.<br><br>Figure out update and delete process. | <b>Status or next steps:</b><br>Working on content of boxes.<br><br>Need to adjust layout and design.<br><br>Need to develop search box. | <b>Status or next steps:</b><br>Will convene in late March/early April.  |

**WHY A BENTO  
SEARCH?**



## Scholarly Articles

### Marine Biology

by Thomas, Robert

*Harvard Review*, 2014, Issue 46, p. 156

### Marine biology: Bacterial boost

by Wake, Bronwyn

*Nature Climate Change*, Jul 1, 2015, Volume 5, Issue 7, p. 621

### Ocean biology: Marine dreams

by Woolston, Chris

*Nature*, Dec 11, 2014, Volume 516, Issue 7530, p. 277

### Teaching "Tropical Marine Biology"

by Landau, M


*INTEGRATIVE AND COMPARATIVE BIOLOGY*, Apr 2015, Volume 55, p. E104

### Marine biology: Charting sea life

by Woolston, Chris

*Nature*, Dec 10, 2015, Volume 528, Issue 7581, p. 295

 See all 527,789 full text online results  More search options

 See all 660,422 results, full text online and citation only

## News & Magazine Articles

### Marine biology Whale fall

by Haag, Amanda

*Nature*, Feb 10, 2005, Volume 433, Issue 7026, p. 566

### Marine biology - Florida red tide brews up drug lead for cystic fibrosis

by Potera, C

*SCIENCE*, Jun 15, 2007, Volume 316, Issue 5831, p. 1561

## Books & eBooks



### Advances in Marine Biology. Volume seventy-four



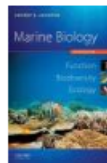
### Marine biology

Castro, Peter

General Collection

QH91 .C37 2010

LOST INVENTORY



### Marine biology : function, biodiversity, ecology

Levinton, Jeffrey S

General Collection

QH91 .L427 2009

LOST INVENTORY



### Web snapshots, 2006- Dept. of Biology and Marine Biology.

University of North Carolina at Wilmington. Dept. of Biology and Marine Biology.

University Archives - Books and Serials

LD3962.9.U5 B565

LIB USE ONLY



### Introduction to marine biology

Karleskint, George

General Collection

QH91 .G37 2006

AVAILABLE

 See all 1,231 results  More search options

## Videos & Music

## Journals, Magazines, & Newspapers by Title

### Marine Biology

### Marine Biology New York

### Marine Biology Research

 See all 3 results  More search options

## Library Websites & Digital Collections

### Biology & Marine Biology | UNCW Randall

Department & Unit Liaison: Biology & **Marine Biology** · Chemistry and Biochemistry · Computer Science · Environmental Studies · Geography and Geology.

[library.uncw.edu/category/unit\\_liaison/biology\\_marine\\_biology](http://library.uncw.edu/category/unit_liaison/biology_marine_biology)



### Peter Fritzler | UNCW Randall

Dec 16, 2015 ... Biology & **Marine Biology**. Chemistry and Biochemistry. Computer Science. Environmental Studies. Geography and Geology. Marine Sciences.

[library.uncw.edu/staff/peter\\_fritzler](http://library.uncw.edu/staff/peter_fritzler)

### Download Original File

establishing the European **Marine Biology** Society. The idea was for marine scientists from both the Soviet bloc and the U. S.-led bloc to meet annually, with the ...

[randall3.uncw.edu/ascod/?p=digitallibrary/getfile&id=245](http://randall3.uncw.edu/ascod/?p=digitallibrary/getfile&id=245)

### Interview with James F. Merritt, November

Nov 9, 2006 ... The biology department had a BS in **marine biology**. That was part of the thing that was started as this marine science program was started

# WHY BENTO?

## For users

- Helps guide & educate users to type of library materials needed
- Good segue into advanced searching
- One search box experience, without overwhelming results
- No facets (at first)

## For Library

- Lack of confidence in discovery, especially for library holdings
- Local catalog data and relevancy retained
- Control over search results
- Summon API robust
- Shortcut to discovery implementation

# WHY NOT BENTO?

- Slow, especially when doing multiple searches
- Not ideal for power users
- “More results” pushes users to different interfaces
- Loss of features such as spell correction, thesaurus, “did you mean”
- Lots of information on the screen
- Hidden collections
- Confusion over where data is coming from



# **BUILDING YOUR BENTO**

# BUILDING YOUR LUNCH



# BUILDING YOUR LUNCH

- How many boxes?
- What to put in each box?
- How to organize boxes?
- How many results to display in each box?
- What source of data to use?
- Does your source data let you accomplish what you want?
- How do you get the data from your source?
- How to design the boxes?

# BUILD ON OTHERS

[search](#)[How to use this search](#)[Articles](#) [Books & Media](#) [Our Website](#) [Research Guides](#) [Images](#) [Other Resources](#)

## Articles

[SEE ALL »](#)

From journals, magazines & conferences

### Marine Biology

by Thomas, Robert

Harvard Review, 03/2017, Issue 46, p. 156

[Journal Article: Full Text Online](#)

### Marine biology: Bacterial boost

by Bronwyn Wake

Nature Climate Change, 07/2015, Volume 5, Issue 7, p. 621

[Journal Article: Full Text Online](#)

### Ocean biology: Marine dreams

## Books & Media

[SEE ALL »](#)

Books, mu

### Oceano Volume

Internet r

### Marine biodive

by Levint

Book, [20:

[Available](#)

### Marine biodive

## Articles



### Marine Biology

by Robert Thomas

Harvard Review, 2014, Issue 46, p. 156

Journal Article

[FIND IT](#)[Report Broken Link](#)

### Marine biology: Bacterial boost

by Bronwyn Wake

Nature Climate Change, 2015-07-01, Volume 5, Issue 7, p. 621

Journal Article

[FIND IT](#)[Report Broken Link](#)

### Ocean biology: Marine dreams

by Chris Woolston

Nature, 2014-12-10, Volume 516, Issue 7530, pp. 277 - 279

Journal Article

[FIND IT](#)[Report Broken Link](#)

## Our Website

[SEE ALL »](#)[Search](#)[Articles](#)[Books & More](#)[Journals](#)[Databases](#)[Guides](#)[Digital Collections](#)[Background Materials](#)[Library Website](#)[Other Searches](#)

## Books & More



### Marine biology: new function, biodiversity, ecology

by Jeffrey S.

BOOKS

Jump to results for: [Scholarly Articles](#) • [News & Magazine Articles](#) • [Ebooks](#) • [Videos](#) • [Databases by Title](#) • [Journals, Magazines, & Newspapers by Title](#)

Oceanogra  
Vol. 36 (Or  
EBOOKS

Oceanogra  
Volume 39  
EBOOKS

Results of  
the Seychel

### Best bet

### Marine biology

### Scholarly Articles

#### Marine Biology (2014)

by Thomas, Robert  
from Harvard Review, issue 46

#### Ocean biology: Marine dreams (2014)

by Woolston, Chris  
from Nature, volume 516, issue 7530

#### Marine biology: Charting sea life (2015)

by Woolston, Chris  
from Nature, volume 528, issue 7581

#### Marine Biology, Old but Unknown (2014)

by Yee-Meng, Jintine Zhu

## Need Help with Your Research?

[Search](#)

### News & Magazine Articles

#### Marine biology Whale fall (2005)

by Haag, Amanda  
from Nature, volume 433, issue 7026

#### Flea market; Marine biology (2013)

from The Economist, volume 406, issue 8823

#### Finding Nemo's role; Marine biology (2013)

from The Economist, volume 406, issue 8825

#### Scents and sensibility; Marine biology (2012)

from The Economist, volume 404, issue 8804

#### Marine biology web (2006)

by Kohn, A.J.  
from CHOICE: Current Reviews for Academic Libraries, volume 44, issue 2

[See more News & Magazine Article results »](#)

### Ebooks

#### Marine Biology: Aquatic Ecology (2011)

by Browne, Sheila A  
published by Nova Science Publishers

#### Marine Biology: Fish Ecology (2013)

by Dempsey, Sean P  
published by Nova Science Publishers

#### Imaging Marine Life: Macrophotography and Microscopy Approaches for Marine Biology (2012)

by Reynaud, Emmanuel G  
published by Wiley-Blackwell

#### What's inside the ocean? (1998)

by Kosek, Jane Kelly  
published by PowerKids Press

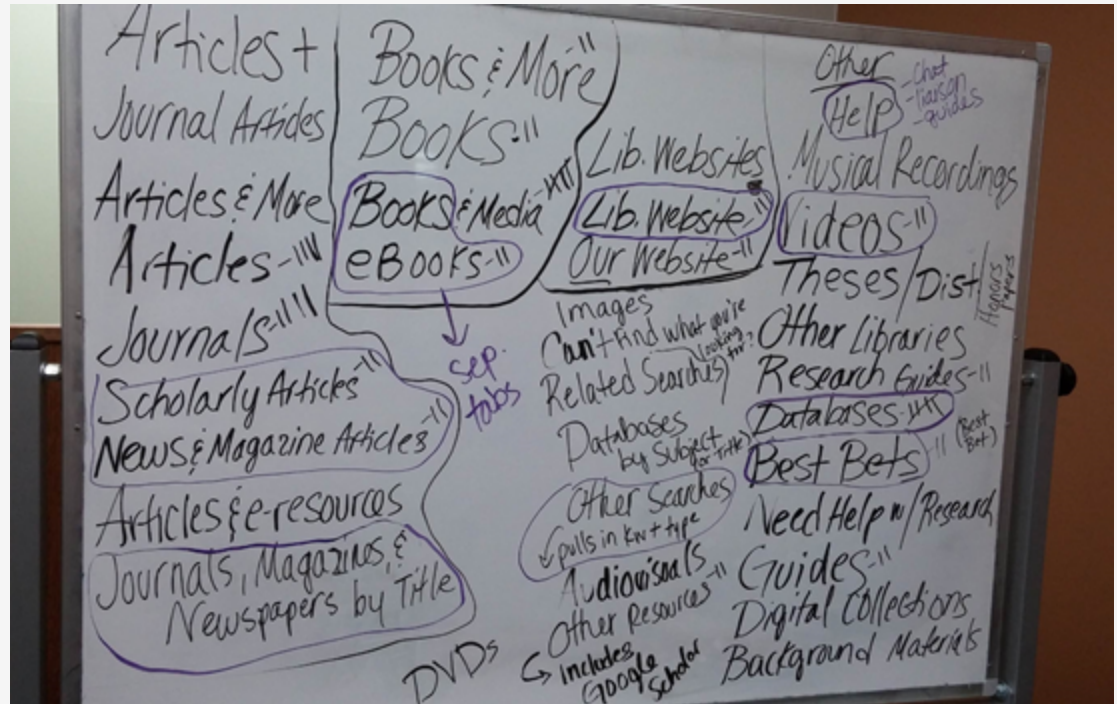
#### Gulf Stream Chronicles: A New Perspective on the Gulf Stream (2014)

by Lee, David S  
published by The University of North Carolina Press



# OUR LUNCH CHOICE

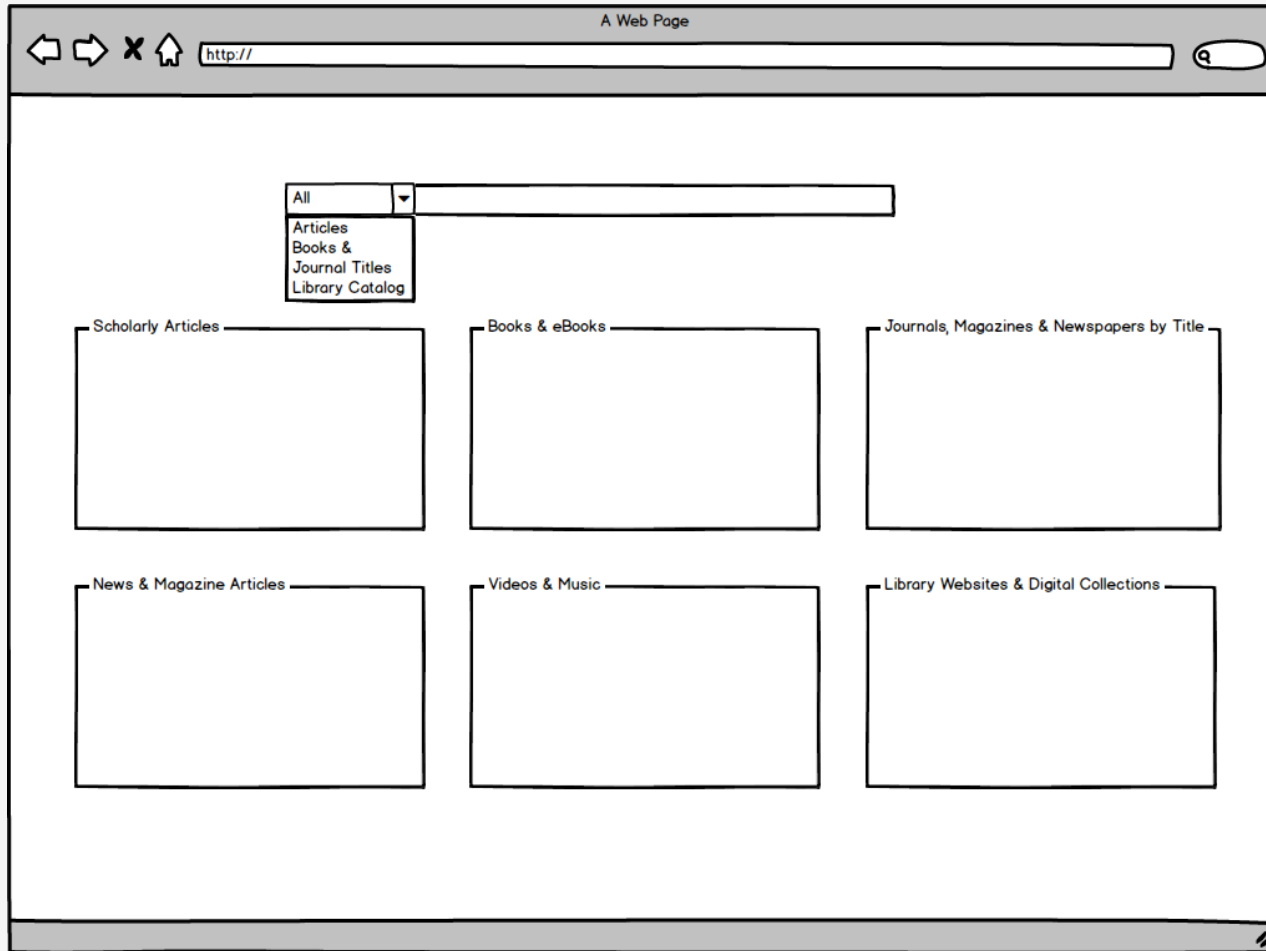
- Scholarly Articles
- News & Magazine Articles
- Databases
- Books & eBooks
- Videos & Music
- Journals, Magazines, & Newspapers by Title
- Library Websites & Digital Collections
- Government Documents
- More Library Collections



# YOUR LUNCH CHOICES

*What categories would you add or remove?*

# ORGANIZING YOUR LUNCH



Wireframing

# SCHOLARLY ARTICLES

- Source: Summon
- Technique: Summon API
- Scope: Full-text Online, Scholarly & Peer-Review, Journal Articles

## Scholarly Articles

### Marine Biology

by Thomas, Robert

*Harvard Review*, 2014, Issue 46, p. 156

### Marine biology: Bacterial boost

by Wake, Bronwyn

*Nature Climate Change*, Jul 1, 2015, Volume 5, Issue 7, p. 621

### Ocean biology: Marine dreams

by Woolston, Chris

*Nature*, Dec 11, 2014, Volume 516, Issue 7530, p. 277

### Teaching "Tropical Marine Biology"

by Landau, M

*INTEGRATIVE AND COMPARATIVE BIOLOGY*, Apr 2015, Volume 55, p. E104

### Marine biology: Charting sea life

by Woolston, Chris

*Nature*, Dec 10, 2015, Volume 528, Issue 7581, p. 295



[See all 527,789 full text online results](#)



[More search options](#)



[See all 660,422 results, full text online and citation only](#)



# NEWS & MAGAZINE ARTICLES

- Source: Summon
- Technique: Summon API
- Scope: Full text online, Newspaper Article, Magazine Articles

## News & Magazine Articles

### Marine biology Whale fall

by Haag, Amanda

*Nature*, Feb 10, 2005, Volume 433, Issue 7026, p. 566

### Marine biology - Florida red tide brews up drug lead for cystic fibrosis

by Potera, C

*SCIENCE*, Jun 15, 2007, Volume 316, Issue 5831, p. 1561

### Flea market; Marine biology



*The Economist*, Feb 16, 2013, Volume 406, Issue 8823, p. 79

### Finding Nemo's role; Marine biology

*The Economist*, Mar 2, 2013, Volume 406, Issue 8825, p. 77

### Scents and sensibility; Marine biology

*The Economist*, Sep 29, 2012, Volume 404, Issue 8804, p. 87

 See all 135,162 full text online results  More search options

 See all 136,186 results, full text online and citation only

# BOOKS & EBOOKS

- Source: Library Catalog (Sierra)
- Technique: screen scraping
- Scope: Limited to specific material types and specific locations
- Book covers: Syndetics

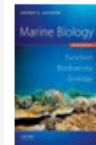
## Books & eBooks



Advances in Marine Biology. Volume seventy-four



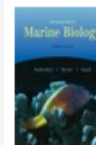
Marine biology  
Castro, Peter  
General Collection  
QH91 .C37 2010  
LOST INVENTORY



Marine biology : function, biodiversity, ecology  
Levinton, Jeffrey S  
General Collection  
QH91 .L427 2009  
LOST INVENTORY



Web snapshots, 2006- Dept. of Biology and Marine Biology.  
University of North Carolina at Wilmington. Dept. of Biology and  
Marine Biology.  
University Archives - Books and Serials  
LD3962.9.U5 B565  
LIB USE ONLY



Introduction to marine biology  
Karleskint, George  
General Collection  
QH91 .G37 2006  
AVAILABLE

 See all 1,231 results  More search options

# VIDEOS & MUSIC

- Source: Library Catalog (Sierra)
- Technique: screen scraping
- Scope: Limited to specific material types and specific locations
- Item covers: Syndetics

## Videos & Music



[Tell Me How Career Series: Marine Biologist](#)



[One ocean Episode 1, Birth of an ocean](#)

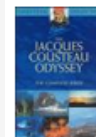


[A sea change](#)



[Treasures of the Great Barrier Reef](#)

Documentary DVDs  
QH197 .T742 2006  
LOST INVENTORY



[The Jacques Cousteau odyssey the complete series](#)

Documentary DVDs  
GC85 .J34 2005 v.1  
AVAILABLE

Documentary DVDs  
GC85 .J34 2005 v.2  
AVAILABLE

[See additional items](#)



[See all 16 results](#)



[More search options](#)

# JOURNALS, MAGAZINES, & NEWSPAPERS BY TITLE

- Source: Library Catalog (Sierra)
- Technique: screen scraping
- Scope: Browse search, 222 (our current Journal Title search)

## Journals, Magazines, & Newspapers by Title

Marine Biology

Marine Biology New York

Marine Biology Research



See all 3 results



More search options

# GOVERNMENT DOCUMENTS

- Source: Library Catalog (Sierra)
- Technique: screen scraping
- Scope: Limited to specific material types and specific locations

## Government Documents

### Graduate studies in marine biology and biological sciences

NC Documents

G130 14:M33

AVAILABLE

### Species profiles: Life histories and environmental requirements of coastal vertebrates and invertebrates Pacific Ocean Region. report 4, the Hawaiian anchovy or nehu, *encrasicholina purpurea* (engrauli)

Clarke, Thomas A

US Government Documents

D 103.24/2:EL-89-10-REPORT 4

AVAILABLE

### Sea turtle research techniques manual

### Minerals Management Service environmental studies program a history of biological investigations in the Gulf of Mexico, 1973-2000

Avent, R. M

### World ocean database 2001

Government Documents Office

C 55.54:OC 2/ 01-02/CD

AVAILABLE

Government Documents Office

C 55.54:OC 2/ 03-04/CD

AVAILABLE

See additional items



See all 99 results



More search options

# LIBRARY WEBSITES & DIGITAL COLLECTIONS

- Source: Library websites and digital collections
- Technique: Google Custom Search
- Scope: Library domains

## Library Websites & Digital Collections

### [Biology & Marine Biology | UNCW Randall Library](#)

Department & Unit Liaison: **Biology & Marine Biology** · Chemistry and Biochemistry · Computer Science · Environmental Studies · Geography and Geology.  
[library.uncw.edu/category/unit\\_liaison/biology\\_marine\\_biology](http://library.uncw.edu/category/unit_liaison/biology_marine_biology)



### [Peter Fritzler | UNCW Randall Library](#)

Dec 16, 2015 ... **Biology & Marine Biology**. Chemistry and Biochemistry. Computer Science. Environmental Studies. Geography and Geology. Marine Sciences.  
[library.uncw.edu/staff/peter\\_fritzler](http://library.uncw.edu/staff/peter_fritzler)

### [Download Original File](#)

establishing the European **Marine Biology** Society. The idea was for marine scientists from both the Soviet bloc and the U. S.-led bloc to meet annually, with the ...  
[randall3.uncw.edu/ascod/?p=digitallibrary/getfile&id=245](http://randall3.uncw.edu/ascod/?p=digitallibrary/getfile&id=245)

### [Interview with James F. Merritt, November 9, 2006 | UNCW Archives ...](#)


Nov 9, 2006 ... The biology department had a BS in **marine biology**. That was part of the thing that was started as this marine science program was started.  
[randall3.uncw.edu/ascod/?p=digitallibrary/digitalcontent&id=86](http://randall3.uncw.edu/ascod/?p=digitallibrary/digitalcontent&id=86)

### [UNCW - Randall Library - Archives](#)

Jan 15, 2009 ... Arbogast, Brian Scott (UNCW, CAS, Department of Biology and **Marine Biology**) Argenbright, Robert (UNCW, CAS, Department of Geography ...  
<https://library.uncw.edu/web/collections/archives/scholarship.html>

[1](#) [2](#) [3](#) [4](#) [5](#) [6](#) [7](#) [8](#) [9](#) [10](#)

powered by Google Custom Search

 [More search options](#)

# DATABASES

- Source: Drupal search
- Technique: Screen scraping
- Scope: Limited by database content type

## Databases

[ASFA - Aquatic Sciences and Fisheries Abstracts](#)

[Oceanic Abstracts](#)

[SciFinder](#)



[See all results](#)



[More search options](#)

# **TECHNICAL SOLUTIONS**



# API



<https://youtu.be/s7wmiS2mSXY>

# USING THE SUMMON API

```
function bento_scholarly_articles() {  
  if (empty($_POST['searchTerms'])) {  
    module_invoke_all('exit');  
    exit;  
  }  
  
  $search_terms = $_POST['searchTerms'];  
  
  // Add filters to the Summon database query.  
  $query = $search_terms . "&s.fvf=IsFullText,true,f&s.fvf=ContentType,Journal+Article,f&s.fvf=IsScholarly,true,f&s.role=authenticated";  
  
  // Query the Summon database and parse the results.  
  $offset = NULL;  
  $limit = NULL;  
  $sort = FALSE;  
  $doctype = 'xml';  
  $api_id = getCredentials('SUMMON__api_id');  
  $api_key = getCredentials('SUMMON__api_key');  
  $xml_summon_data = summon_tools__request($api_id, $api_key, $query, $offset, $limit, $sort, $doctype);  
  $parsed_summon_data = summon_tools__parse_xml("summon_articles", $xml_summon_data, $exact);  
  
  // Return the results.  
  $results = $parsed_summon_data["summon_articles"]["results"];  
  $results_count = $parsed_summon_data["count"];  
  $results_returned = array();  
  $results_returned["results"] = $results;  
  $results_returned["resultsCount"] = $results_count;  
  drupal_json_output($results_returned);  
  module_invoke_all('exit');  
  exit;  
}
```

# USING THE SUMMON API

```
function summon_tools__request($api_id, $api_key, $query, $offset = 0, $limit = 10, $sort = False, $doctype = "xml") {
    /* Returns response in $doctype format (xml|json) from Summon 2.0 API for a $query.

    - Results start at $offset for a total of $limit results.
    - Results are sorted by relevance ($sort = False) or date-descending ($sort = True).
    - Code based on Python code here: https://gist.github.com/lawlesst/1070641
    - See also: http://blog.humaneguitarist.org/2014/09/04/getting-started-with-the-summon-api-and-python/
    - See also: http://blog.humaneguitarist.org/2014/11/16/using-the-summon-api-with-a-simple-php-script/
    */

    // set API host and path.
    $host = "api.summon.serialssolutions.com";
    $path = "/2.0.0/search";

    // create query string.
    $query = "s.q=" . $query . "&s.pn=$offset&s.ps=$limit&s.ho=true";

    // set sort to date-descending if needed.
    if ($sort != False) {
        $query = $query . "&s.sort=PublicationDate:desc";
    }

    // sort and encode $query.
    $query_sorted = explode("&", $query);
    asort($query_sorted);
    $query_sorted = implode("&", $query_sorted);
    $query_encoded = urldecode($query_sorted);

    // create request headers.
    $accept = "application/$doctype";
    $date = gmstrftime("%a, %d %b %Y %H:%M:%S GMT", time());
    $id_string = implode("\n", array($accept, $date, $host, $path, $query_encoded, ""));
    $digest = base64_encode(hash_hmac("sha1", utf8_encode($id_string), $api_key, True));
    $authorization = "Summon " . $api_id . " " . $digest;
    $headers = array("Host:$host", "Accept:$accept", "x-summon-date:$date", "Authorization:$authorization");

    // call API; return response.
    $url = "http://$host$path?$query";
    $response = url_tools__request($url, $timeout = 10, $headers = $headers);
    return $response;
}
```

# DISPLAYING THE RESULTS

```
success: function(response) {
    var display = '';
    var results = response.results;
    var limit = results.length;
    if (limit > 5) {
        limit = 5;
    }
    if (limit) {
        var i = 0;
        var result;
        do {
            result = results[i];
            display += '<div class="bento-result">';
            // Journal title.
            if (result.title && result.link) {
                display += '<a href="' + result.link + '" target="_blank"><b>' +
                    result.title + '</b></a><br>';
            }
            else if (result.title) {
                display += '<b>' + result.title + '</b><br>';
            }
            // Author.
            if (result.author) {
                display += 'by ' + result.author + '<br>';
            }
            // Source.
            if (result.randall_source) {
                display += result.randall_source + '<br>';
            }
            display += '</div>';
            i++;
        } while (i < limit);
    }
```

# SCREEN SCRAPING



[http://www.iharvest.co.uk/wp-content/uploads/2014/03/iHarvest\\_Screen\\_Scrape\\_Image\\_Of\\_Text.png](http://www.iharvest.co.uk/wp-content/uploads/2014/03/iHarvest_Screen_Scrape_Image_Of_Text.png)

# PHP SIMPLE HTML DOM PARSER

- An HTML DOM parser written in PHP5+ that lets you manipulate HTML
- It finds tags on an HTML page by using selectors, like [jQuery](#). (the JavaScript library)
- It extracts contents from HTML and returns it in a variable that can then be displayed.

# HOW SCREEN SCRAPING?


```
foreach ($records as $record) {
    if (++$j == ($results_count + 1)) {
        break;
    }
    $detail = array();
    $detail['record_additional_items'] = FALSE;
    $detail['record_url'] = $record->getElementsByTagName('a', 0)->href;
    $detail['record_title'] = $record->getElementsByTagName('a', 0)->plaintext;
    $detail['record_elements'] = $record->find('.bibItemsEntry td');
    $detail['record_elements_count'] = count($detail['record_elements']);
    if ($detail['record_elements_count'] > 0) {
        $data = $record->getElementsByTagName('td', 0)->plaintext;
        $detail['record_collection'] = _bento_clean_string($data);
        $data = $record->getElementsByTagName('td a', 0)->href;
        $detail['record_collection_url'] = _bento_clean_string($data);
        $data = $record->getElementsByTagName('td', 1)->plaintext;
        $detail['record_call_number'] = _bento_clean_string($data);
        $data = $record->getElementsByTagName('td a', 1)->href;
        $detail['record_call_number_url'] = _bento_clean_string($data);
        $data = $record->getElementsByTagName('td', 2)->plaintext;
        $detail['record_availability'] = _bento_clean_string($data);
    }
}
```

# SCREEN SCRAPING



of 25)

Most relevant titles entries 1-2




The girl on the train  
Hawkins, Paula, author  
New York : Riverhead Books, 2016  
★★★★☆  
**LOCATION**      **CALL #**      **STATUS**

|                                    |                                      |              |
|------------------------------------|--------------------------------------|--------------|
| <a href="#">General Collection</a> | <a href="#">PR6108.A963 G57 2016</a> | DUE 05-04-17 |
|------------------------------------|--------------------------------------|--------------|

```
</th>
<th width="24%" class="bibItemsHeader">
STATUS
</th>
</tr>
<tr class="bibItemsEntry">

<td width="33%" ><!-- field 1 -->&nbsp;&nbsp;&a
href="/screens/floor_gen_coll.html">General Collection</a>
</td>
<td width="43%" ><!-- field C -->&nbsp;&nbsp;&a
href="/search=S4?/cPR6108.A963+G57+2016/cpr+6108+a963+g57+2016/-3,-1,,E/browse">PR6108.A963 G57 2016</a> <!-- field v -->
<!-- field # -->&nbsp;&nbsp;&</td>
<td width="24%" ><!-- field % -->&nbsp;&nbsp;&DUE 05-04-17 </td>
</tr>
</table>
</div>
```



The girl on the train  
Hawkins, Paula, author  
[General Collection](#)  
[PR6108 A963 G57 2016](#)  
[DUE 05-04-17](#)



# **DIFFICULTIES OF SCREEN SCRAPING**

- You are relying on the consistency of the display on another site
- Different scenarios have to be handled differently
- It can be really difficult to find a way to map to the data you desire

# SCREEN SCRAPING ISSUES

- Multiple Results

```
<div class="briefcitDetailMain">

<!--{nohitmsg}-->
<h2 class="briefcitTitle">
<a
href="/search~S4?/Xgirl+on+a+train&SORT=DZ/Xgirl+on+a+train&SORT=DZ&extended=0
t&FF=Xgirl+on+a+train&SORT=DZ&7%2C7%2C">In a dark, dark wood</a></h2>
<br >
Ware, Ruth<br >
New York : Scout Press, 2015
  <!--{year}-->
```

- One Result

```
564 <!-- BEGIN BIBDETAIL TABLE BORDER -->
565 <table id='bib_detail' width="100%" border="0" cellspacing="1" cellpadding="2" class="bibDetail">
566 <tr class="bibInfoEntry" ><td>
567 <!-- BEGIN INNER BIB TABLE -->
568 <table width="100%" cellspacing="3" cellpadding="0">
569 <tr><!-- next row for fieldtag=t -->
570 <td valign="top" width="20%" class="bibInfoLabel">Title</td>
571 <td class="bibInfoData">
572 <strong>In <span style="color:RED" ><strong>a</strong></span> dark, dark wood / Ruth Ware</strong></td></tr>
573 </table>
574 <!-- END INNER BIB TABLE -->
575 </td></tr>
576 </table>
577 <!-- END BIBDETAIL TABLE -->
578 </h1>
```

# NO RESULTS

## Books & eBooks

No results found in the UNCW library catalog.

Additional results from other libraries:

**How to build a girl : a novel**

Moran, Caitlin, and Louise Brealey

2014. Internet resource

[Request Item through Interlibrary Loan](#)

[Check New Hanover County Public Library or Cape Fear Community College.](#)

**How to Build a Girl.**

*How to Build a Girl*

Print.

[Request Item through Interlibrary Loan](#)

[Check New Hanover County Public Library or Cape Fear Community College.](#)

# WORLD CAT API

```
function _bento_oclc_search_api($search_terms, $category, $worldCatSearchType = 'equal') {

    $url_parameters = '';

    if ($category == 'books') {
        //$formats = array('print', 'sound+recording', 'internet+resource');
        $url_parameters = 'q=srw.dt+exact+%22bks%22+or+srw.dt+exact+%22url%22+and+srw.ti=%22';
    }
    elseif ($category == 'videos') {
        $url_parameters = 'q=srw.dt+exact+%22vis%22+or+srw.dt+exact+%22sco%22+or+srw.dt+exact+%22rec%22+and+srw.ti+all+%22';
    }
    elseif ($category == 'journals') {
        //$formats = array();
    }

    //key to the OCLC API|
    $wskey = getCredentials('OCLC_wskey');
    $search_terms = urlencode($search_terms);
    $search_terms = str_replace('%22', '', $search_terms);
    $search_terms = str_replace('"', '', $search_terms);
    $display_part1 = '<b>No results found in the UNCW library catalog. <br /> Additional results from other libraries:
</b><br /> <br />';
    $display_part2 = '';

    // Limit results to 2 matches.
    $oclc_matches = 0;

    // Call the OCLC Search API with our search terms.
    $oclcurl = 'http://www.worldcat.org/webservices/catalog/search/worldcat/opensearch?' . $url_parameters . $search_terms
        . '%22&format=rss&wskey=' . $wskey . '&cformat=mla';

    // Use the MagpieRSS parser to turn the reply from OCLC into an object.
    $rss = fetch_rss($oclcurl);

    // Take a look at each record returned by OCLC.
    foreach ($rss->items as $item) {
        if ($oclc_matches >= 2) {
            break;
        }

        $holdings = FALSE;
        $isbn = '';
        $isbn1 = '';
        $oclc = '';
        $requested_format = FALSE;
```

# GOOGLE CUSTOM SEARCH

- Problem: How to search our many websites and digital collections in a unified fashion?
- Solution: Google Custom Search <https://developers.google.com/custom-search/>

# GOOGLE CUSTOM SEARCH



Search in CSE home



## Custom Search

New search engine

▼ Edit search engine

Search Randall Lib

Setup

Look and feel

Search features

Statistics and Logs

Business

► Help

Send Feedback

Basics

Make money

Admin

Advanced

Provide basic details and preferences for your search engine. [Learn more](#)

### Search engine name

Search Randall Library's Websites

### Search engine description

Search across Randall Library websites

### Search engine keywords

Search engine keywords, e.g. climate 'global warming' 'greenhouse gases'

### Edition

Nonprofit, ads optional.

### Details

Search engine ID

Public URL

Get code

### Image search

ON

### Speech Input

ON

### Language

English

Advanced

### Sites to search

Search only included sites

Add

Delete

Filter

Label

1- 4 of 4



☐ Site

Label

☐ <http://archivesspace.uncw.edu/>

☐ <http://digitalcollections.uncw.edu/cdm/>

☐ <http://randall3.uncw.edu/assignmentc...>

☐ [library.uncw.edu](http://library.uncw.edu)

marine biology

Search

x

Demo search box

Web Image

About 80 results (0.35 seconds)

Sort by: R

### [Biology & Marine Biology | UNCW Randall Library](#)

Department & Unit Liaison: Biology & Marine Biology · Chemistry and Biochemistry · Computer Science · Environmental Studies · Geography and Geology.

[library.uncw.edu/category/unit\\_liaison/biology\\_marine\\_biology](http://library.uncw.edu/category/unit_liaison/biology_marine_biology)



### [Peter Fritzler | UNCW Randall Library](#)

Dec 16, 2015 ... Biology & Marine Biology. Chemistry and Biochemistry. Computer Science. Environmental Studies. Geography and Geology. Marine Sciences.  
[library.uncw.edu/staff/peter\\_fritzler](http://library.uncw.edu/staff/peter_fritzler)

### [UNCW - Randall Library - Archives](#)

Jan 15, 2009 ... Arbogast, Brian Scott (UNCW, CAS, Department of Biology and Marine Biology) Argenbright, Robert (UNCW, Department of Geography ...

<https://library.uncw.edu/web/collections/archives/scholarship.html>



### [UNCW: Randall Library - University Archives - Buildings & Landmarks](#)

Known as Dobo Hall since its 1999 dedication to the Dobo brothers, it is currently used by the Department of Biology and Marine Biology and the Department of ...



<https://library.uncw.edu/web/collections/archives/bnl/25.html>

### [Find Theses, Dissertations, Honors Papers | UNCW Randall Library](#)

Oct 23, 2014 ... UNCW Theses, Dissertations, Honors Papers and Special Projects are searchable through the Library Catalog, author name, title or ...

[library.uncw.edu/theses\\_dissertations](http://library.uncw.edu/theses_dissertations)

# GOOGLE CUSTOM SEARCH



## Custom Search

New search engine

▼ Edit search engine

Search Randall Lib

Setup

Look and feel

Search features

Statistics and Logs

Business

▼ Help

Help Center

Help forum

Support

Blog

Documentation

Terms of Service

Send Feedback

### Get code for search results

Copy the following code, and paste it into a <div> element in your site's <body> section, where you want the search results to render.

**Note:** For the most cross-browser compatibility, it is recommended that your HTML pages use a supported doctype such as <!DOCTYPE html>. CSS hover effects require a supported doctype.

```
<script>
(function() {
  var cx = '005978018638547358306:xrqdgqxe2na';
  var gcse = document.createElement('script');
  gcse.type = 'text/javascript';
  gcse.async = true;
  gcse.src = 'https://cse.google.com/cse.js?cx=' + cx;
  var s = document.getElementsByTagName('script')[0];
  s.parentNode.insertBefore(gcse, s);
})();
</script>
<gcse:searchresults-only></gcse:searchresults-only>
```

You can customize the Search UI even more, or add per page customization by following the [full documentation on CSE element](#).

Search Results Details

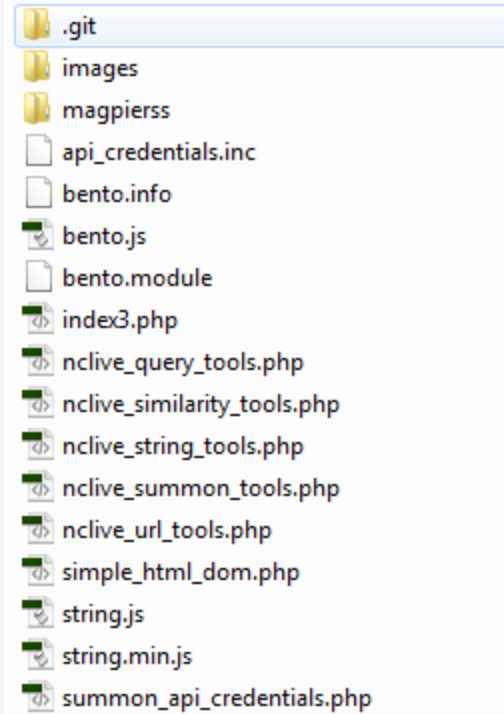
See HTML example

© 2017 C

# DRUPAL: MODULE

- Pulls all of the above together into Drupal
- Uses Drupal functions to run queries, store and display the data
- Two main elements bento.module and bento.js
- Other tools to parse the XML returned from the APIs
- Available on Github

*\*Thanks to NCLIVE for sharing their Drupal code & Summon API tools.*





# DRUPAL: JS

- Takes search terms entered by user and queries all the data sources.
- Shows and hides the appropriate parts of the page. Parts are hidden until the application has obtained data to put in those containers.
- Some results can quickly begin to appear on the page.
- Sends data to bento.module to parse for display.

## Scholarly Articles



## News & Magazine Articles



## Databases



# DRUPAL DISPLAY: FRAME

```
// COLUMN 2
//
'<div class="bento-results-col bento-results-col2">

  <div id ="bento-results-books" class="bento-results">
    <div class="bento-results-header">
      <h5>Books & eBooks</h5>
    </div>
    <div class="bento-results-content bento-results-content-books bento-results-columns bento-dynamic">
      
    </div>
    <div class="bento-results-footer">
      <img class="bento-see-all"/>
      <a class="bento-see-all-link" href="#" target="_blank"></a>
      <img class="bento-more-options"/>
      <a class="bento-more-options-link" href="#" target="_blank"></a>
    </div>
  </div>

  <div id ="bento-results-videos" class="bento-results">
    <div class="bento-results-header">
      <h5>Videos & Music</h5>
    </div>
    <div class="bento-results-content bento-results-content-videos bento-results-columns bento-dynamic">
      
    </div>
    <div class="bento-results-footer">
      <img class="bento-see-all"/>
      <a class="bento-see-all-link" href="#" target="_blank"></a>
      <img class="bento-more-options"/>
      <a class="bento-more-options-link" href="#" target="_blank"></a>
    </div>
  </div>
'</div>' .
```

# DRUPAL DISPLAY: INNARDS

```
// BOOKS

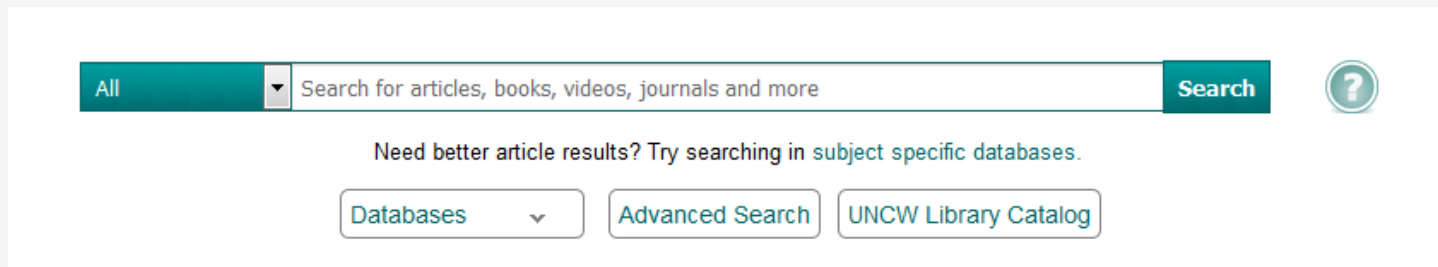
$('.bento-results-content-books').html(loading);
var booksUrl = systemPointer + '/bento_catalog_books';

$.ajax({
  type: 'POST',
  url: booksUrl,
  dataType: 'json',
  data: {'searchTerms': searchTerms},
  success: function(response) {
    var results = $.trim(response.reply);
    $('.bento-results-content-books').html(results);
    if (results === 'No results found') {
      var differentSearch = 'http://libcat.uncw.edu/search/X?SEARCH=(' + encodeURIComponent(searchTerms) + ')&searchscope=4';
      var worldcatSearch = 'http://www.worldcat.org/search?q=' + encodeURIComponent(searchTerms);
      var nhcplSearch = 'http://srvlibweb.nhcgov.com/?config=pac#section=search&term=' + encodeURIComponent(searchTerms);
      var cfccSearch = 'http://encore.cfcc.edu/iii/encore_cfcc/search?lang=eng&target=' + encodeURIComponent(searchTerms);
      var alternativeSearches = '. Try a different search in the <a href="' + differentSearch;
      alternativeSearches += '" target="_blank"> UNCW Library Catalog</a>.<br><br>';
      alternativeSearches += 'Or search <a href="' + worldcatSearch + '" target="_blank">WorldCat.org</a>, ';
      alternativeSearches += '<a href="' + nhcplSearch + '" target="_blank">New Hanover County Public Library</a>, ';
      alternativeSearches += 'or <a href="' + cfccSearch + '" target="_blank">Cape Fear Community College</a>.<br><br>';
      alternativeSearches += 'Can\'t find what you need? <a href="/ask" target="_blank">Ask Us!</a>';
      $('.bento-results-content-books').append(alternativeSearches);
    }
    else {
      var viewAll = 'http://libcat.uncw.edu/search/X?SEARCH=(' + searchTerms + ')&searchscope=4&SORT=D&m=a&m=c&m=h&b=wg&b=wj&b=wr&b=wf&b=wh&b=wb&b=wc&b=we&b=wi&b=wl&b=wn&';
      if ((response.resultsCount !== '1') && (response.resultsCount !== 'X')) {
        $('#bento-results-books .bento-results-footer img.bento-see-all').attr("src", externalLink);
        $('#bento-results-books .bento-results-footer img.bento-more-options').attr("src", moreOptions);
        $('#bento-results-books .bento-results-footer img').show();
        $('#bento-results-books .bento-results-footer a.bento-see-all-link').attr("href", viewAll);
        $('#bento-results-books .bento-results-footer a.bento-see-all-link').text("See all " + numberWithCommas(response.resultsCount) + " results");
        $('#bento-results-books .bento-results-footer a.bento-more-options-link').attr("href", catalogAdvancedPart1 + searchTerms + catalogAdvancedPart2 + searchTerms);
        $('#bento-results-books .bento-results-footer a.bento-more-options-link').text("More search options");
        $('#bento-results-books .bento-results-footer a').show();
      }
    }
  },
  error: function() {
    $('.bento-results-content-databases').html('<span class="bento-error">This search of the Library Catalog could not be completed.</span>');
  }
});
```

# **IMPLICATIONS & FUTURE**

# IMPLICATIONS

- Changes to bib codes (e.g., eBooks) to be able to limit search
- Affected Summon implementation choices (e.g., MARC mapping)
- Homepage search box



The screenshot shows the top section of the UNCW Library Catalog homepage. It features a search bar with a teal dropdown menu set to 'All'. The search bar contains the placeholder text 'Search for articles, books, videos, journals and more'. To the right of the search bar is a teal 'Search' button and a circular help icon with a question mark. Below the search bar, there is a text prompt: 'Need better article results? Try searching in [subject specific databases](#).' Underneath this prompt are three buttons: 'Databases' with a dropdown arrow, 'Advanced Search', and 'UNCW Library Catalog'.

All ▼ Search for articles, books, videos, journals and more Search ?

Need better article results? Try searching in [subject specific databases](#).

Databases ▼ Advanced Search UNCW Library Catalog

# FUTURE CONSIDERATIONS

- Response time
- Move away from screen scraping
- Missing collections
- More usability testing
- EBooks
- Audiobooks
- Alternatives like Blacklight or VuFind

**QUESTIONS?**

# IMAGE CREDITS

- <https://www.tastemade.com/articles/8-throwback-facts-about-lunchables-to-make-you-super-nostalgic>
- <http://www.molevol.org/the-buffet-and-the-sushi-conveyor/>
- <https://www.youtube.com/watch?v=7TvVxPzyGTY>