WEB FIRST

PASSION. INNOVATION. SOLUTIONS.

Drupal and the PubMed API

Connecting a Medical Publications Library to PubMed:

Case Study of NIA's Alzheimer's Preclinical Trial Database (AlzPED)

Introduction / Speakers

Sanjay Patel, AlzPED Project Manager, WebFirst

Ted Gifford, Sr. Developer, WebFirst

James King, Branch Chief and Information Architect, NIH Library

WebFirst – Areas of Expertise

- Federal Agency & nonprofit clients
- Web Design and Information Architecture
- Content Management Systems
  - Drupal, WordPress
  - Portals, Collaboration, eStores
- Social Media – Twitter, Facebook, LinkedIn, YouTube, RSS Feeds, Tag Clouds
- Mobile Technologies
  - Android / iOS
  - Data Collection
- Accessibility/Section 508 Compliance
- Informatics / Data Visualization
- Geographic Information Systems (GIS)

Hackpad

https://drupalgovcon2015.hackpad.com/

About the AlzPED website

- Alzheimer's Disease Preclinical Efficacy Database
- Purpose: Promote pre-clinical therapy development for AD
- Knowledge platform
  - Information dissemination
  - Promote efficiency, transparency, reproducibility
- Includes published and unpublished studies
- Anticipated September launch

About the AlzPED website

- Some implementation details
  - Drupal 7
  - Apache Solr Integration
    - Facet API, Autocomplete, Views Integration
  - Multistep forms for data entry
  - PubMed API Integration

What is PubMed?

- You may have heard of it
- > 24 million citations for biomedical literature
- Links to full-text versions of citations
- A critically important tool for researchers
  - Broad coverage of topics and publications
  - Search tools are powerful
- “Industry Standard”
  - Used in third-party tools
  - Developed and maintained by NCBI @ NLM @ NIH
- One of the Entrez databases

How Researchers use PubMed

- Find publications related to their research
- Import results into citation management software
- Searches can use MeSH terms (taxonomy)
  - "Alzheimer Disease/therapy"[MH] => 17921 results
- Most searches start with free-text these days
  - alzheimer's disease therapy => 30976 results

Can I get some Drupal with that?

Entrez Programming Utilities

- E-Utilities, API
- Covers 38 databases, including PubMed
- Nine endpoints (with “E” prefix)
  - Info (database statistics)
  - Search (text searches)
  - Post (UID uploads/Batch API)
  - Summary (document summaries/teaser)
  - Fetch (data record downloads)
  - Link (Entrez links)
  - GQuery (global query)
  - Spell (spelling suggestions)
  - CitMatch (batch citation searching)

Entrez Programming Utilities

- E-Utilities, API
- Covers 38 databases, including PubMed
- Nine endpoints (with "E" prefix)
  - Info (database statistics)
  - **Search (text searches)**
  - Post (UID uploads/Batch API)
  - **Summary (document summaries/teaser)**
  - Fetch (data record downloads)
  - **Link (Entrez links)**
  - GQuery (global query)
  - Spell (spelling suggestions)
  - CitMatch (batch citation searching)

Example API Request: Search

- http://eutils.ncbi.nlm.nih.gov/entrez/eutils/esearch.fcgi?db=pubmed&term=cancer&reldate=60&datetype=edat&retmax=2
- Response:


```
▼<eSearchResult>
  <Count>21935</Count>
  <RetMax>2</RetMax>
  <RetStart>0</RetStart>
 ▼<IdList>
    <Id>26193703</Id>
    <Id>26193700</Id>
  </IdList>
 ▼<TranslationSet>
  ▼<Translation>
     <From>cancer</From>
    ▼<To>
       "neoplasms"[MeSH Terms] OR "neoplasms"[All Fields] OR "cancer"[All Fields]
     </To>
   </Translation>
  </TranslationSet>
 ▼<TranslationStack>
  ▼<TermSet>
     <Term>"neoplasms"[MeSH Terms]</Term>
     <Field>MeSH Terms</Field>
     <Count>2686660</Count>
     <Explode>Y</Explode>
   </TermSet>
  ▼<TermSet>
     <Term>"neoplasms"[All Fields]</Term>
     <Field>All Fields</Field>
     <Count>2127895</Count>
     <Explode>N</Explode>
```

Example API Request: Search (Raw)

- http://eutils.ncbi.nlm.nih.gov/entrez/eutils/esearch.fcgi?db=pubmed&term=cancer&reldate=60&datetype=edat&retmax=2
- Response:


```
<?xml version="1.0" encoding="UTF-8"?>
<!DOCTYPE eSearchResult PUBLIC "-//NLM//DTD esearch 20060628//EN"
"http://eutils.ncbi.nlm.nih.gov/eutils/dtd/20060628/esearch.dtd">
<eSearchResult><Count>22229</Count><RetMax>2</RetMax><RetStart>0</
RetStart><IdList><Id>26204561</Id>
<Id>26204558</Id>
</IdList><TranslationSet><Translation>    <From>cancer</From>
<To>"neoplasms"[MeSH Terms] OR "neoplasms"[All Fields] OR "cancer"[All
Fields]</To>  </Translation></TranslationSet><TranslationStack>  <TermSet>
<Term>"neoplasms"[MeSH Terms]</Term>  <Field>MeSH Terms</Field>
<Count>2688128</Count>  <Explode>Y</Explode>  </TermSet>  <TermSet>
<Term>"neoplasms"[All Fields]</Term>  <Field>All Fields</Field>
<Count>2129138</Count>  <Explode>N</Explode>  </TermSet>  <OP>OR</
OP>  <TermSet>  <Term>"cancer"[All Fields]</Term>  <Field>All Fields</
Field>  <Count>1574665</Count>  <Explode>N</Explode>  </TermSet>
<OP>OR</OP>  <OP>GROUP</OP>  <TermSet>  <Term>2015/0
```

Example API Request: Summary

- http://eutils.ncbi.nlm.nih.gov/entrez/eutils/esummary.fcgi?db=pubmed&retmode=json&id=11850928,11482001
- Response:


```
{
    "header": {
        "type": "esummary",
        "version": "0.3"
    },
    "result": {
        "uids": [
            "11850928",
            "11482001"
        ],
        "11850928": {
            "uid": "11850928",
            "pubdate": "1965 Aug",
            "epubdate": "",
            "source": "Arch Dermatol",
            "authors": [
                {
                    "name": "LoPresti PJ",
                    "authtype": "Author",
                    "clusterid": ""
                },
                {
                    "name": "Hambrick GW Jr",
                    "authtype": "Author",
                    "clusterid": ""
                }
            ],
            "lastauthor": "Hambrick GW Jr",
            "title": "Zirconium granuloma following treatment of rhus dermatitis.",
            "sorttitle": "zirconium granuloma following treatment of rhus dermatitis ",
```

Example API Request: Link

- http://eutils.ncbi.nlm.nih.gov/entrez/eutils/elink.fcgi?dbfrom=pubmed&db=pubmed&id=20210808&cmd=neighbor_score
- Response:

```
▼<eLinkResult>
  ▼<LinkSet>
    <DbFrom>pubmed</DbFrom>
    ▼<IdList>
      <Id>20210808</Id>
    </IdList>
    ▼<LinkSetDb>
      <DbTo>pubmed</DbTo>
      <LinkName>pubmed_pubmed</LinkName>
      ▼<Link>
        <Id>15876306</Id>
        <Score>75133399</Score>
      </Link>
      ▼<Link>
        <Id>15466689</Id>
        <Score>66865452</Score>
      </Link>
      ▼<Link>
        <Id>12956738</Id>
        <Score>64959303</Score>
      </Link>
      ▼<Link>
        <Id>11149990</Id>
        <Score>61522749</Score>
      </Link>
      ▼<Link>
        <Id>9275137</Id>
        <Score>61364505</Score>
      </Link>
```

A Note on Appropriate API Usage

- Behave
  - *NCBI recommends that users post no more than three URL requests per second and limit large jobs to either weekends or between 9:00 PM and 5:00 AM Eastern time during weekdays.*
  - Failure to comply with this policy may result in an IP address being blocked
  - Register your tool ID and email address
- Drupal considerations
  - Cache data retrieved from the API
  - Make tool ID and email address configurable

Use Cases for Drupal Integration

Make life easier for your users

- Don't tell them "*Entrez, maintenant partez*"

Better Search

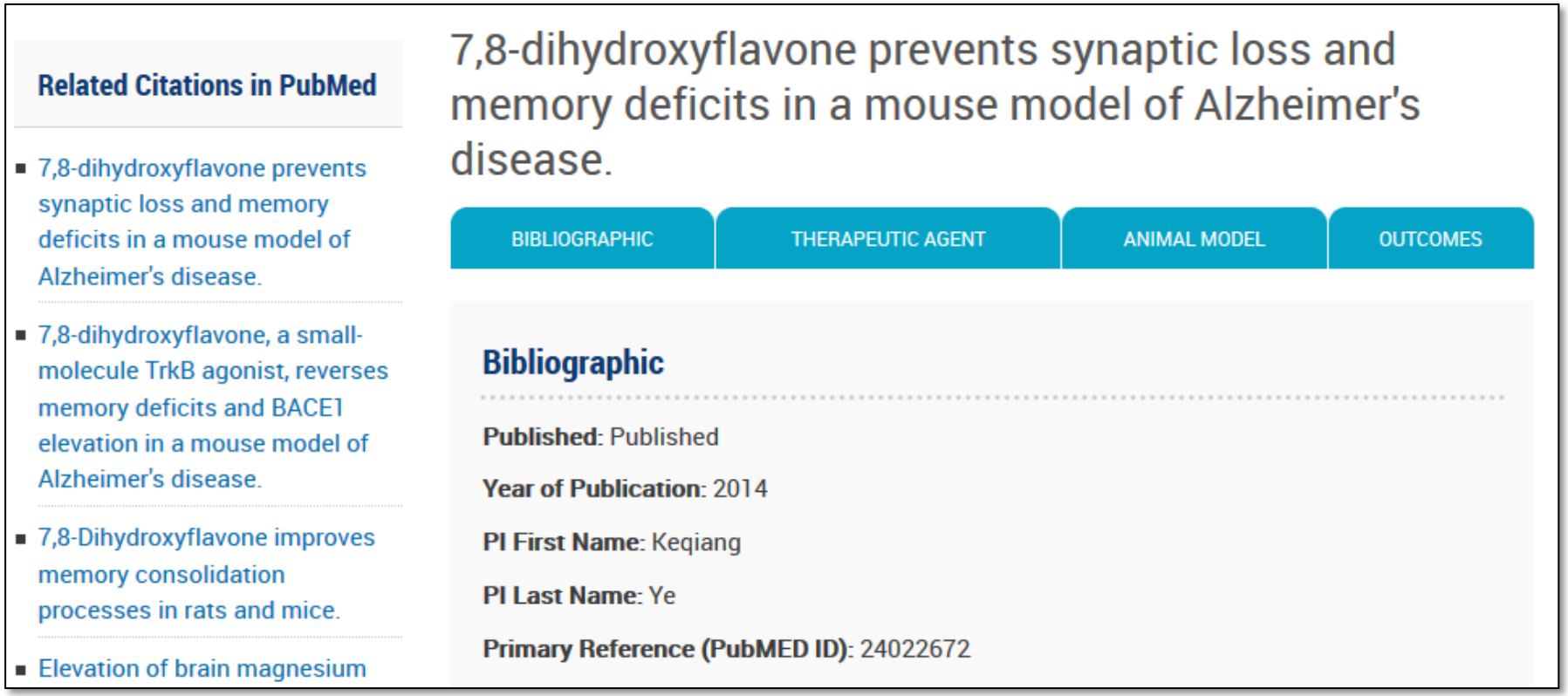
Data Entry/Data Quality

Use Cases: Citation/Resource Library

- For a team, or the general public
- Import citations directly from PubMed
- Aid in searching for new literature
- Export to desktop citation manager software

Use Cases: Show Related Citations

- Link to PubMed, or
- Link to the citations on your Drupal site



Use Cases: Data Entry / Submission

- Pre-fill the form if the user has a PubMed ID
- Check for spelling errors or other data quality issues

Use Cases: Taxonomy Synchronization

- Synchronize a Drupal taxonomy with MeSH terms
- Don’t re-invent the ontological round thingey

Use Cases: Make Autocomplete Smarter

- Get spelling suggestions
- Search for matching MeSH terms
- …but be careful not to overload the API

Use Cases: EMR-Literature Link

- Making relevant clinical data available to physicians
  - In-context
  - With institution-specific content or commentary
  - Link to PubMed and additional resources

Use Cases: Third-Party Services

- Integrate with
  - PubReminer
  - Chilibot
  - Etc.

Apps

Use Cases: Apps

- Your app can integrate directly with the PubMed API, or
- Integrate indirectly via the Drupal API you are already developing for the app.

“There’s a module for that.”

Existing Drupal Modules

- That use the PubMed API
- Searched for
  - “pubmed” and “entrez” on d.o, including sandbox projects
  - “pubmed drupal” on GitHub
  - …and similar searches on Google
- Found five modules
  - None provide a generic API library
  - There is significant overlap in features

Bibliography Module (biblio)

- Drupal 6.x-7.x
- Includes support for importing from PubMed based on PMID
  - You wouldn't know it from the project page
- Related Module: Entrez Database Import (entrez)
  - Drupal 6.x
  - Import to biblio by PMID **or search term**
  - Import via Batch API or cron
  - Views and Flags for moderation of imported citations
- Related Module: Pubmed Integration (pubmed_integration)
  - Drupal 6.x
  - Similar features

PubMed Field (sandbox)

- Drupal 7.x
- Provides a field that can be attached to any fieldable entity
- Displays list of citations matching query specified in field
- Cron integration
- Themeable

MeSH Auto-Import (sandbox)

- Drupal 7.x
- Retrieves MeSH terms by PMID
- Assigns terms to hard-coded term reference field, which must be created manually

AlzPED Related Citations

- Drupal 7.x
- Uses ELink Utility instead of ESearch
- Related citations stored in a multi-value link field (with title)
- Ordered by similarity to AlzPED resource

Web Service Client(s) Modules

- Abstract the details of an API/web service into configuration
  - No longer need to maintain an Entrez-specific web service client within the preceding modules
- May not be worth the extra complexity
- Several modules that perform similar functions
  - Web Service Client (wsclient)
  - Web Service Clients (clients)
  - Services Client (services_client)
  - Comparison: https://www.drupal.org/node/2292623

Other PHP Code (non-Drupal)

- Four Entrez API clients on GitHub
  - None cover the entire API
  - Can offer inspiration

Next Steps

- Work with other maintainers to create a base API client library
- Consider consolidating existing modules
- Evaluate Web Service Client(s) modules
- Start work on a D8 module

Questions

Resources

- Entrez Programming Utilities: http://www.ncbi.nlm.nih.gov/books/NBK25501/
- Drupal Modules:
  - Bibliography: https://www.drupal.org/project/biblio
  - Pubmed Integration: https://www.drupal.org/project/pubmed_integration
  - Entrez Database Import: https://www.drupal.org/project/entrez
  - PubMed Field: https://www.drupal.org/sandbox/nellessen/1780320
  - MeSH Auto-Import: https://www.drupal.org/sandbox/smithdalec/1593502
  - Web Service Client https://www.drupal.org/project/wsclient
  - Web Service Clients https://www.drupal.org/project/clients
  - Services Client https://www.drupal.org/project/services_client
  - WS Client Comparison https://www.drupal.org/node/2292623
- Non-Drupal PHP Code:
  - https://github.com/UMNLibraries/ncbi-eutils-client-php
  - https://github.com/asifr/PHP-PubMed-API-Wrapper
  - https://github.com/tmpjr/pubmed
  - https://github.com/jfmbrennan/php-pubmed-api/blob/master/libs/PubMedApi.php

WebFirst, Inc.
15800 Crabbs Branch Way, Suite 120
Rockville, MD 20855

(301) 670-1690

info@WebFirst.com

www.WebFirst.com