LIBRARY SCIENCE
Integrating Drupal with DAMS
A Quick Introduction

Who am I?
Aaron Bauman d.o user: aaronbauman
Why Integrate?

- DAMS are usually powerful but ugly.
- Expose material in an interactive, curated user experience.
- You can appeal to a wider audience, by providing your content in multiple forms.
- Ease the content management burden when new material moves around on its own.
What do I need?

- Someone who understands how to get stuff out of your DAMS.
- Someone who understands how to import things into Drupal.

Both of these are usually harder than they first appear.
Case Studies

Two examples:
Beinecke Library at Yale
Historical Society of Pennsylvania
→ Replace old static HTML site.
→ Several DAMS that needed to be discoverable in one place.
→ Provide rich media experience for great material.
→ Ease the flow of new material from their DAMSs to the site.
→ Keep the integration light weight.
Beinecke Library

Level 1: Integrated Searches
Beinecke Library

Level 2: XML and RSS import

<table>
<thead>
<tr>
<th>Import</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>New Acquisitions</td>
<td>Imports the list of new acquisitions in the system.</td>
</tr>
<tr>
<td>New Finding Aids</td>
<td>Import new finding aids.</td>
</tr>
</tbody>
</table>
Level 3: Relate content and export
Digital History Exhibits Project
Client need: Branded digital exhibit template
Drupal content generated by importing a TEI-encoded XML document
Pulls digital assets from DAMS via API
Exhibit Structure

→ Objectives
  ◆ Drupal admins can create new exhibits as needed
  ◆ Curated narrative to expose depth of content
  ◆ Relate content to primary asset (documents)

→ Application
  ◆ Node type drives context via Reference fields
  ◆ Feeds, XML Parser
  ◆ Customizations to accommodate TEI, field collections, other exigencies
TEI Parser

→ Objectives
  ◆ Semantic encoding of historical documents
  ◆ Consolidation and unification: D.R.Y.

→ Application
  ◆ Import TEI (xml) file to create/update Drupal nodes
    ● Feeds
    ● XPath Parser
    ● hooks
    ● patches
  ◆ Warning: feeds and field collections don’t get along
Historical Society of Pennsylvania

Notable Hooks and patches

- views_query_alter
- block_view_alter
- menu_block_tree_alter
- search_api_page_search_form_alter
- search_api_query_alter
- menu_alter
- formfeeds_import_form_alter
- feeds_after_parse
- feeds_presave
- Patches created by Message Agency are always contributed back to drupal.org
- 1424086-ignore_destination_GET_parameter-1.patch
- 2006870-1-negated_fulltext_filters.patch
- date-date_plain_format_settings-1806856-2.patch
- date-plain_text_formatter_respect_settings-1763788-3.patch
- eck-admin_paths-1836892-1.patch
- entityreference-feeds_integration-1616680-66.patch
- feeds-DateTime_support_for_php_5.3-1258582.patch
- field_collection-feeds-1063434-93.patch
- field_collection_feeds_multiple.patch
- hide-empty-field-collections-1276258-33_0.patch
- leaflet-markercluterer-outsideInteraction-100644.diff
- relation-7.x-1.x-dev--feeds-integration-1345958.patch
- search_api-1694832-7-field_changes.patch
- search_api-add_templates-1313676-26.patch
- search_api_db-prevent_infinite_recursion-1917592_0.patch
- search_api_facetapi_0.patch
- timeline-date-field.patch
Historical Society of Pennsylvania

Closed for Business
The Story of Bankers Trust Company During the Great Depression

Import | Description
--- | ---
Multi-Document Corpus | MULTIPLE file upload for documents containing multiple transcripts

Genres
Organizations
People
Topics
Historical Society of Pennsylvania
Historical Society of Pennsylvania

**DOCUMENT DETAILS:**

**Source Image**
- 90:5284-958.64.1402.15.1

**Document Title**
- Telegram from Harland Buckman to Albert Greenfield

**DATE (SEARCH)**
- [ ] Show End Date
  - [ ] Show End Date
  - End Date: 1830/12/22
  - Format: 2013/10/13

**User-Facing Date**
- December 22, 1830

**AUTHOR(S):**
- [ ] Add another item
  - [ ] Harland Mcknight Buckman (4027)

**RECIPIENT(S):**
- [ ] Add another item
  - [ ] Albert Monroe Greenfield (4203)
XSL Filter

→ Objectives
  ◆ Non-destructive import of TEI markup
  ◆ Provide source XML documents for download

→ Application
  ◆ TEICHI input filter provides starting point
  ◆ Use XSL to transform TEI markup into HTML
XSL Filter

→ Key features

◆ Tailored XSLT and XSL stylesheet
◆ “Fuzzy dates” for display; real dates for search
◆ XML id -> Feeds GUID -> Node id translation
◆ Modals!
PATRONS ARE REQUESTED TO FAVOR THE COMPANY BY CRITICISM AND SUGGESTION CONCERNING ITS SERVICE.

This is a full-rate Telegram or Cablegram unless its deferred character is indicated by a suitable sign above or preceding the address.

WESTERN UNION

NEWCOMB CARLTON

Text format: TEI

Source Information

Text format: TEI

Source Page

Source ID

5284-1959_b14002_1_TEI.xml

Remove
### Document Details:

**Source Image**

90:5284-1936_B14002_1.MF

**Document Title**

Telegram from Irland Beckman to Albert Greenfield

**DATE (SEARCH)**

- **Show End Date**
  - 1920/12/22
  - Format: 2013/10/11

**User-Facing Date**

December 22, 1920

**AUTHOR(S):**

- Irland McKnight Beckman (4027)

**RECIPIENT(S):**

- Albert Monroe Greenfield (4253)
Document Viewer

→ Objectives
◆ Expose DAMS assets to Drupal, via TEI
◆ Page and transcript side-by-side
◆ Access to high resolution image for pan and zoom

→ Application
◆ Collective Access API
◆ Asynchronous image caching
◆ New York Times Viewer (OSS, via github)
**Document Details:**

<table>
<thead>
<tr>
<th>Source Image *</th>
<th>90:5284-1958_614002_1.tif</th>
</tr>
</thead>
<tbody>
<tr>
<td>Document Title *</td>
<td>Telegram from Inland Beckman to Albert Greenfield</td>
</tr>
</tbody>
</table>

**Date (Search):**
- Show End Date
- 1920/12/22
- Format: 2013/10/23

**User-Facing Date:**
- December 22, 1920

**Author(s):**
- Inland McNikey Beckman (4027)
- [Add another item](#)

**Recipient(s):**
- Albert Monroe Greenfield (4253)
- [Add another item](#)
Telegram from Irland Beckman to Albert Greenfield, December 22, 1930

Telegram from Irland Beckman to Albert Greenfield

Author(s): Irland McKnight Beckman  |  Recipient(s): Albert Monroe Greenfield
December 22, 1930

PATRONS ARE REQUESTED TO FAVOR THE COMPANY BY CRITICISM AND SUGGESTION CONCERNING ITS SERVICE

CLASS OF SERVICE

This is a flat-rate telegram of Cablegram unless its deferred character is indicated by a suitable sign above or preceding the address.

WESTERN UNION

NEWCOMB CARLTON, PRESIDENT

J. C. WILLEVER, FIRST VICE-PRESIDENT
Historical Society of Pennsylvania

Source Information:
Title
Telegram from Irland Beckman to Albert Greenfield

Note 1
This meeting no doubt will recall that Bankers Trust Co. had turned over its affairs of Banking.
Albert Monroe Greenfield

Albert M. Greenfield was a prominent Philadelphia businessman involved in real estate, banking, transportation, and retail, as well as philanthropy, civic affairs, and politics. He had
PHILADELPHIA PENN DEC 22 400A 1930 ALBERT M GREENFIELD DELIVER IMMEDIATELY GET RECEIPT 350 WEST

THIS MORNING DECEMBER 22ND AT EIGHT O'CLOCK YOUR PRESENCE IMPERATIVE

[short text]

...ARE APPROPRIATE GIFTS FOR ALL OCCASIONS.

WESTERN UNION

ARE REQUESTED TO FAVOR THE COMPANY BY CRITICISM AND SUGGESTION CONCERNING ITS SERVICE

SIGNS

D.L. = Day Letter
N.M. = Night Message
N.L. = Night Letter
D.C. = Deferred Cable
N.L.T. = Cable Night Letter
W.L.T. = Week-End Letter

Event Timeline and Map

→ Objectives
  ◆ Personal narratives of the Underground Railroad
  ◆ Extract normalized data from historic documents
  ◆ Expose stories via spatial relationships
  ◆ Navigate spatiotemporal data

→ Application
  ◆ Geocode location data
  ◆ Views-driven, contextual map + timeline
  ◆ Time Mapper: http://timemapper.okfnlabs.org
Historical Society of Pennsylvania

Locations & Events + Timeline + Map = 
Historical Society of Pennsylvania

Map

Timeline
circa late 1855 - early 1856:

Canada

The Vigilance Committee in Philadelphia intends to send Harriet Shephard's group to Canada, but divides and disguises the group for safety.

Related: Anna Maria Shephard, Edwin Shephard, Eliza Jane Shephard, Harriet Shephard, John Henry Shephard, Mary Ann Shephard, Unnamed Fugitive Man 1, Unnamed Fugitive Woman 2, Hiram Wilson
Relationship Network

→ Objectives
◆ Surface relationships between all individuals
◆ Discover buried stories within network
◆ Visualize interactions between key players

→ Application
◆ D3.js Force Directed network graph [http://d3js.org](http://d3js.org)
◆ Again, views-driven, contextual graph
◆ MIT Media Lab “Immersion” project: [https://immersion.media.mit.edu/demo](https://immersion.media.mit.edu/demo)
Historical Society of Pennsylvania
Lessons Learned

- The data will not compile itself
- Planning is important
- Flexibility is more important
- The data may not actually tell a good story
- Limits of Drupal contrib
- Decide on a vendor, not a pricetag
The data will not compile itself

- Massive volume of man hours parsing historic documents, organizing and normalizing.

- Underestimating or underfunding this effort will impact a project timeline.

- Once the data is extracted, make sure you have a good place to put it.
Planning is important

- What data is worth collecting?
- How granular can the data be?
- What DAMS suits our needs?
- What goals should the application fulfill?
- What should be prioritized, and what can slip?
- What is the Minimum Viable Product?

- Bring vendors, consultants, collaborators in early.
- But not too early.
Flexibility is more important

- Planning the data schema is necessary to integrate with 3rd party systems, but some degree of schema shift should be expected.
- Plan an extensible, pluggable system.
- The lowest bidder is never planning for the next project.
- Don’t hack core.
- Don’t hack contrib.
- Do not reinvent the wheel.
Data may not actually tell a good story

- You may be expecting too much from your data.
- Data curation can be just as effort intensive as data collection.
- Data by itself is not necessarily interesting.
- Think about your target audiences and user stories.
- Design and presentation are just as important as content in order to reach your constituents.
Limits of Drupal contrib

- Please refer to slide titled “Notable Hooks and patches”
- Just because there’s a module called “Disqus” doesn’t mean it does what you think it does.
- Just because you know how two modules work doesn’t mean you know how they work together.
- Conflicting pressures: “Right Way” vs. “Fast Way”
Decide on a vendor, not a pricetag

- Estimation and RFPs are guess work.
- Specify budget or features. Not both.
- Having a good relationship with a talented vendor will save you time, money, and headache. Full stop.
Questions?
Resources

- D3.js
  http://d3js.org

- Immersion
  http://immersion.media.mit.edu

- TimeMapper
  http://timemapper.okfnlabs.org

- NYT Viewer
  https://github.com/documentcloud/document-viewer

- Leaflet
  http://leafletjs.com

- TEI

- TEICHI
  http://www.teichi.org/

- Collective Access
  http://www.collectiveaccess.org
Aaron Bauman
aaron@messageagency.com
http://drupal.org/user/384578
Salesforce Suite Project Owner & Maintainer

http://messageagency.com
info@messageagency.com
(215) 546-6496