HAT:

An extension to the Islandora Image Viewer

Table of Contents

- Table of Contents
- Changes to existing software
  - Islandora Drupal module
    - plugins/Newspapers.inc - (modified)
  - Islandora Image Viewer
    - viewer.jsp - (modified)
    - js/iiv.js - (modified)
    - js/iivutil.js - (new)
  - Fedora Service Definitions
    - newspapers:viewerSdep-issueCModel - (modified)
    - newspapers:viewerSdep-pageCModel - (modified)
- New modules
  - iiv_annotation
  - guardian_solr_config
- Scope/Features (Semi-technical)
- Over and above
  - Padding
  - Ubuntu Virtual Machine Image
  - Solr index
  - Polygon annotations
  - Trigger workflow
  - IIV text search
- Architecture Diagrams
- Future Considerations
- Sign-off

Changes to existing software

From the existing Islandora solution stack, modifications occurred within the following pieces:

Islandora Drupal module

plugins/Newspapers.inc - (modified)

This file has been modified so that additional parameters can be passed into the Islandora Image Viewer upon launch. These parameters include the drupal host domain, the drupal user id, and the current solr search term.

Islandora Image Viewer

viewer.jsp - (modified)

This file has been modified to accept the new parameters passed through the Newspaper.inc file in the Drupal module. It sets the current solr search term, the current Drupal user and current Drupal domain as JavaScript variable available within the viewer.

js/iiv.js - (modified)

This file has been modified to include functions for:

- auto searching for the solr search term to perform text highlighting;
- adding the user interface requirements for annotations
    - dropdown,
    - flagging,
    - drawing
- user interface requirements for internal highlighting searches;
- searching for new annotations/text-highlights upon page turn.

js/iivutil.js - (new)

This file has been created by the HAT project team.Inside includes all functions to:

- draw annotations (polygon and rectangular),
- draw highlight boxes,
- AJAX functions for communications against iiv_annotation module,
    - add annotation,
    - receive annotations,
    - search for highlight coordinates, and;
    - flag annotations
- functions to clear the highlight/annotation layer

- function to automatically populate the annotation dropdown.

Fedora Service Definitions

newspapers:viewerSdep-issueCModel - (modified)

Changes were made to the METHODMAP and WSDL data-streams to accommodate the additional parameters being passed to the IIV viewer through Newspaper.inc.

newspapers:viewerSdep-pageCModel - (modified)

Same changes that were performed to the issueCModel.

New modules

iiv_annotation

This module supports all features for text highlighting and annotation creation, retrieval and flagging. The module has been designed with intentions that all functions will be called through AJAX. One important note is that all data returned is in the form of JSONP. The use of JSONP allows cross-domain requests--effectively eliminating all issues with the same origin policy. The module also includes an administration screen to adjust the URL of the Fedora server, adjust the padding on text highlighting boxes, and to monitor flagged annotations. Trigger support has been added for the flagging piece, so that custom work flows may be developed around the event.

guardian_solr_config

This module is a display override for the solr search block. It overwrites the display functionality, so that all results are outputted with the desired search term.

Scope/Features

HAT has two major functions: highlighting and annotations. The new features concerning the highlighting piece are:

- **Search:** A word or phrase can be searched by using the text box in the upper right hand corner. Searched items will appear in a highlight box wherever they appear in the newspaper.

- **Hide Text Boxes:** The user can choose to hide the searched items by clicking the ‘Highlight Text’ button. This button is found at the end of the button list on the top button bar, it looks like a pencil.
- **Change Highlight Size:** The search boxes can be changed in size by entering the Drupal site configuration menu->iiv_annotation_flag menu. On this page there is a small section where vertical and horizontal values can be added or subtracted to the highlight boxes.
- **Zoom Functionality:** The user can change the zoom of the page with the zoom buttons and keep the highlighted search boxes over search items.
- **Page Functionality:** The user may change the newspaper page and keep the search valid for the next page. In other words, the search and highlight text functionality can be used for the whole newspaper archive.

The new features concerning the annotation piece are:

- **Regular Annotations:** A user can create a rectangular annotation by clicking the ‘Draw Rectangle’ button and dragging a rectangle on the canvas. The ‘Draw Rectangle’ button is the second last button on the bottom toolbar.
- **Multi-sided Annotations:** A user can create a muti-sided or polygon annotation by clicking the ‘Draw Polygon’ button and clicking points on the newspaper image. These points will be the vertices of the polygon annotation, where the annotation will fill automatically. The ‘Draw Polygon’ button is the last button on the bottom toolbar.
- **Hide Annotations:** All annotations can hidden like the highlight boxes by clicking the ‘Hide Annotation’ button, which is the first bottom in the second toolbar.
- **Flagging Functionality:** A user can flag an annotation by selecting the annotation and clicking the ‘Flag Annotation’ button. Once an annotation is flagged, it will no longer be available to public users until an administration removes the flag in the Drupal control configuration. This configuration page is the same page with highlight box size functionality: configuration menu->iiv_annotation_flag menu.
- **Public Functionality:** The user can choose to make a annotation public or private on creation of an annotation. To make a annotation public click the public check box in the annotation text area.
- **Comment Functionality:** The comment functionality, is a short message describing the annotation, other than the annotation text. A user can set the comment when he/she first creates the annotation.

Over and above

Padding

When we started using the coordinate data from the WORDSXY files, we came to realize that some of the coordinate were not very accurate. As such, we decided it might be a

good idea to allow the coordinates used when drawing the boxes themselves to be altered slightly. With the assumption that the coordinates are--on average--good, we have implemented the ability to scale the highlighting box around the original coordinates.

To allow an administrator to control the amount of padding used in the text highlighting component, we created an administration page, which permits the adjustment of two Drupal variables. These two variables are used as a percentage/scaling factor to increase (or possibly decrease) the amount padding in both the horizontal and vertical directions.

Ubuntu Virtual Machine Image

To help speed up development, we thought it would be ideal to have each group member run an instance on Fedora/Islandora, so that they could test their code before committing changes to source control--one issue with this approach: The image with which we were originally provided was a 64-bit distribution, which requires specialized hardware in order to run. The solution was simply to move to a 32-bit distribution.

Due to both familiarity and the knowledge that as a distribution it is kept up to date, we built another Virtual Machine Image based on the 32-bit version of Ubuntu Server, and migrated Fedora, Islandora and associated modules to this new image. We found that in addition to being able to be run on a wider range of machines, the Ubuntu image seemed to be both faster and more stable.

Solr index

In order to highlight text in images, we first had to be able to determine which images contained instances of the given query. As such, it was necessary first to have the WORDSXY data stream indexed in some manner. Due to slow communication at the time, we elected to build our own Solr index and request handler for this purpose.

It is not explicitly necessary to have the WORDSXY file indexed in particular--the OCR file might be processed somewhat more efficiently. The main point is that the full-text from each paper page be indexed and associated with the appropriate PID.

The use of Solr to directly facilitate selecting of the text to highlight was also considered but dropped due to time constraints.

Polygon annotations

Non-rectagular polygons can be used for annotations by clicking on the “Draw Polygon” button. The user can then click arbitrarily on the map/image to make the desired shape. The coordinates of these polygons are stored similarly to those for rectangular polygons, using the method *saveAnnotation*. To get the annotations back into IIV requires two steps: First, the coordinates are received as a string. Then, these coordinates are parsed into a form directly usable by OpenLayers.

Polygons are drawn either on the existing annotation layer, or a new layer is created if one is not found. An anchored popup-box with an opacity of 0.8 is populated with text content of the annotation, and assigned the same ID as the polygon feature ID in OpenLayers.

Trigger workflow

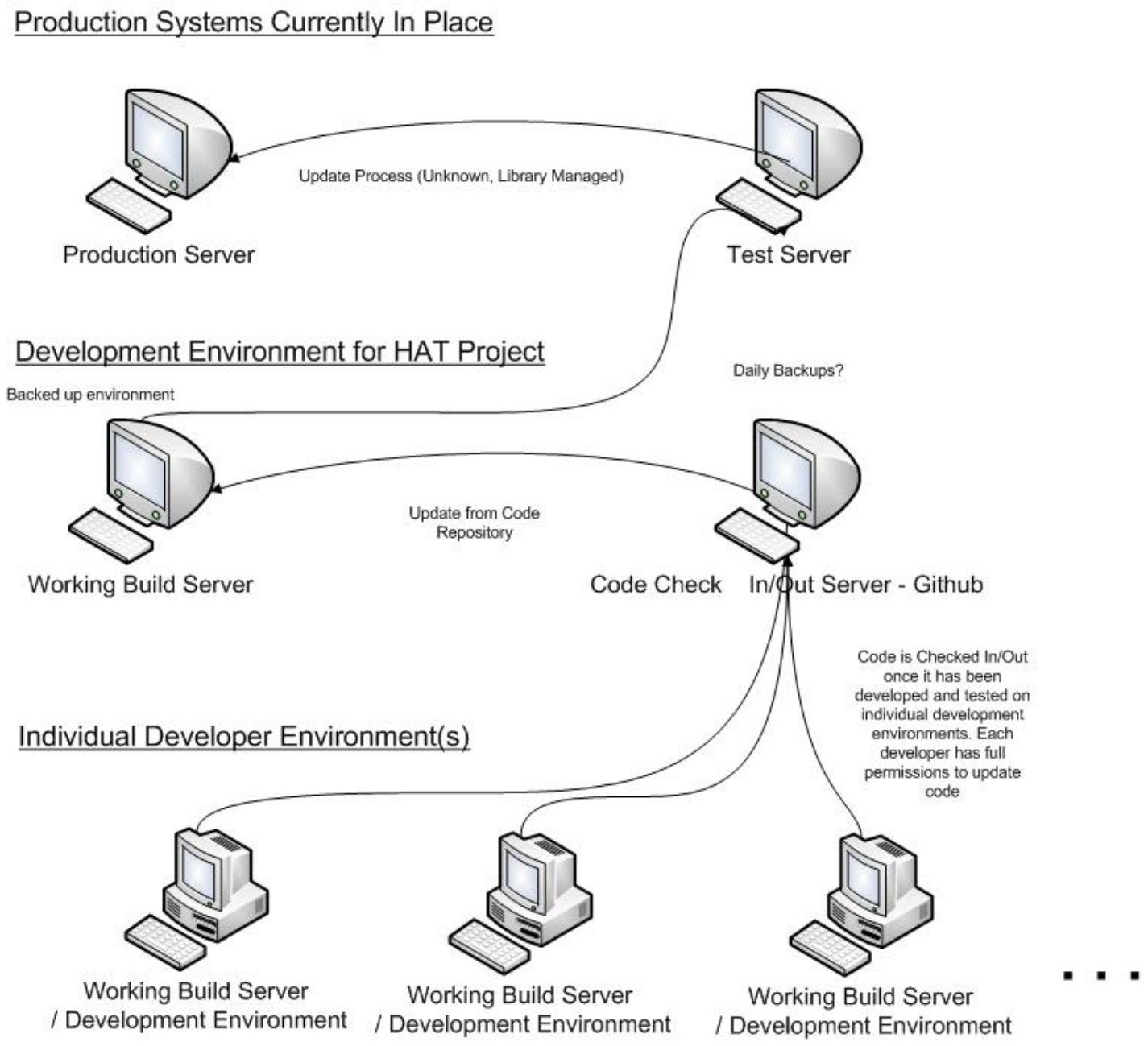
The annotation module has a flagging event which can be tied in with the rules module. This allows custom work flows to be built around the flagging event. Such events could include emailing; watchdog custom log message; and auto blocking a user.

IIV text search

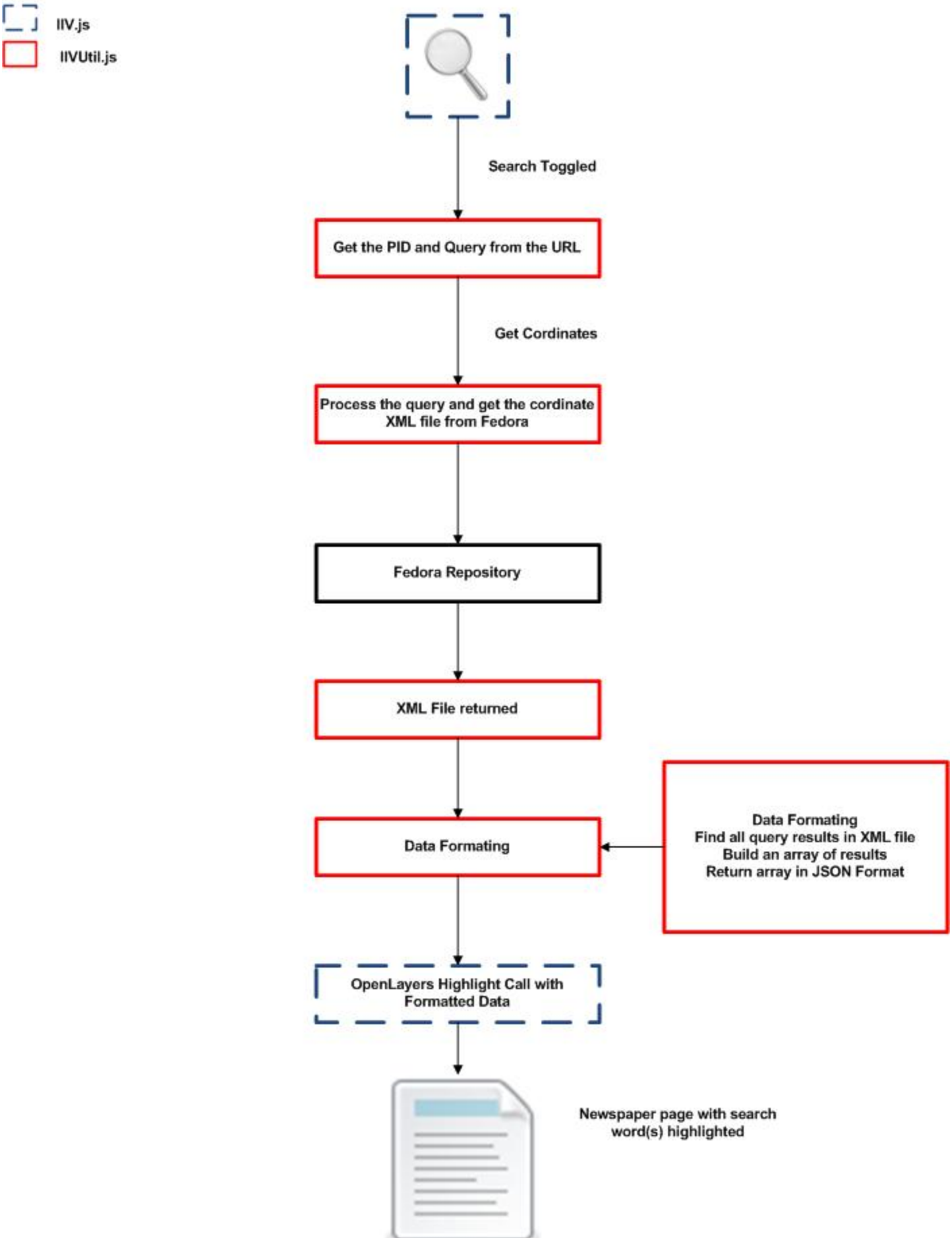
We provide functionality in the IIV_annotation module for using a query to search in the WORDSXY coordinate file, which is stored within the Fedora repository. It supports searching for multiple words and fuzzy search. It returns JSON data, which contains the coordinates of the words which match the query.

Architecture Diagrams

Development Methodology / Setup for HAT project Team



Highlight of Search Words



Add Annotation

Public

Private

Client selects to add an annotation to an image

No
Error Message Returned

Ajax call:
/islandora/annotations/insert/
$pid/$text/$notation/$public

Authorized User

Drupal

User is notified that annotation has been saved for public / private viewing

Private and Public Annotation Storage

Drupal DB

Annotations Table

PID
UID
Notation
PublicFlag

Public Annotation Storage

Fedora

Fedora XML File

PID
Notation

Annotation Status Message

View Annotation

Select annotation(s) on menu bar

No Error Message Returned

Ajax Call: /islandora/annotations/select/ $myPID

Authorized User

Drupal

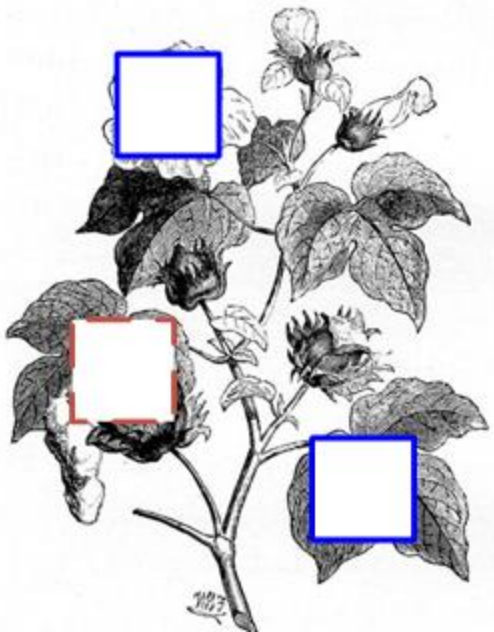
Image presented with relevant annotation overlay

Drupal DB

Return message: Annotation data Public annotations Private annotations with matching UID's

Annotations Table
PID UID Notation PublicFlag

User Table
UID Etc..



Public

Private

Future Considerations

- Fedora authorization to use the roles from Drupal properly (allow annotators to annotate)
    - This is a matter of configuring Fedora
- Determine how the Solr search index relates to the text highlighting search
- Build more informative front end
    - Dither buttons when permissions aren't available
    - Allow users to delete their own annotations from the front end
    - Indicate to users when annotations have already been flagged
- Allow annotations to be edited

HAT Project Sign-off

Purpose:

This document acknowledges that the client have been provided a demo and project documentation outlining the delivered functionality.

Terms:

The client is satisfied will the base functionality delivered and the project obligations have been met.

Software Warranty: The software is provided as-is and is available to the open source community.

Signed by Client	Signed by Professor/Development Team
	Professor:
Donald Moses Dated	Prof. David Leblanc Dated
	Development Team PM:
	Jeremy Noonan Dated