Implementation and Development Call 20130926

Calls are held every Thursday at 1 pm eastern time (GMT–4 in daylight savings, GMT-5 standard time) – convert to your time at http://www.thetimezoneconverter.com

View and edit this page permanently at https://wiki.duraspace.org/x/O-MQAg, or use the temporal Google Doc for collaborative note taking during the call.

VIVO is hiring!

DuraSpace is seeking a dynamic and entrepreneurial Project Director for the open source VIVO project (www.vivoweb.org), a world-wide community focused on creating software tools, ontologies, and services. The VIVO Project Director will have the opportunity to play a major role in a collaborative movement that will shape the future of research.

See full posting

Release update

Hoping to start testing next Monday when Jim returns.

Apps and Tools Group

Notes from Sept. 24 meeting recorded as this webcast.

Next meeting on Tuesday in two weeks (October 8) at 1pm Eastern.

Demonstrated a set of Python tools developed at UF to run a set of SPARQL queries nightly to detect malformed or missing connections, duplicate identifiers, and data that should not be in VIVO due to privacy concerns. Reports come back as plain text that gets emailed out, and reports are structured to return a value of zero when there are no anomalies that need to be addressed. So far it's just a notification tool, and doesn't do the cleanup.

The queries are pretty generic and are intended to be easily modified through a configuration file, or they could be run using a tool like CURL.

The tools are available at http://github.com/nrejack/dchecker and while still being developed are usable already. A demonstration video is available at: http://www.youtube.com/watch?v=8Lz4V7HuETk.

Chris put examples of the Apache rewrite rules in last week's Implementation and Development call notes, along with code used to generate the list of mapping rules from UF Gator IDs to VIVO URIs.

Upcoming Events

- 2nd Annual CASRAI International Conference, October 16-18 in Ottawa
  - Conference streams: Reconnect Big Data, Reconnect the Library, and Reconnect the Machine
  - http://reconnect.casrai.org
  - **Jon will be presenting on VIVO, along with Memorial University**

- 1st Annual UCosmic Conference, October 31 in New York
  - **Collaborative Software Development to Address Strategic Challenges in Higher Education: Kuali, VIVO and UCosmic**
  - http://www.ucosmic.org/Conference.aspx

Updates

- Brown (Ted) - finalizing public rollout schedule, adding data from existing profile system, working on scripts to find new publications for faculty from Web of Science (using the Lite API), Pubmed, Springer Metadata APIs.
  - have been doing a lot of work to get DOIs for publications to help with disambiguation
    - CrossRef OpenURL web service - helpful when you know ISSN, volume or issue, starting page number but not DOI
      - http://labs.crossref.org/openurl/
  - what about publications without DOIs, or where can't find them? At this point, focusing on data with DOIs, including on reviews, a common type of publication for humanities faculty
  - Springer OA is a metadata API covering most things published by Springer
    - http://dev.springer.com/
  - Web of Science Lite web service
    - documentation: http://wokinfo.com/products_tools/products/related/webservices/ws_faq/
- Colorado (Alex) - no major updates on VIVO implementation, working on bringing publications data into Elements for first 1,200 faculty /researchers, about way thru first pass of curation, finding DOIs very useful where they are found, including in some of the Bibtex or RIS data sources like EBSCO, ProQuest, and Google Scholar. Elements has a Bibtex importer but has to be filtered by the individual author and assumes the author is already claimed, rather then a Bibtex file having publications for a large group of authors with the option of having publications only marked as still pending approval by the author. Bringing up a new server for VIVO 1.6.
- Cornell (Brian, Tim, Jon, Huda) -- working on VIVO 1.6, merging in the changes for the VIVO-ISF ontology and grepping the rest of the code for examples of property names that have changed. Adding the ability to pull in Library of Congress Subject Headings (LCSH) with the LOC-assigned URIs.
- Duke (Richard) - Mainly working on some data cleanup tasks with orphaned entities - like publications that aren't linked to anyone.

- EPA (Cristina, Zac) - No updates on going live. Hoping there is no government shutdown next week to delay us. Currently working with the Freemarker SPARQL Data Getters for some custom reports and having a few issues with that (Alumni and Expertise reports). Tried out the example, and we think it works, but we don't have any orgs classified as an academic department. Trying to find where people went to school so can see all the people at EPA who went to a given school, as well as all the people associated with an EPA-defined vocabulary term defined in SKOS vs. terms brought in from UMLS or GEMET (this may be related to the external terms only being typed as owl:Thing). Getting some duplicate values that are attributable. The actual queries work, but when we try to make a custom template to better format the information (.ftl) it fails.
- Florida (Nicholas) -- see update above from the Apps & Tools meeting
- RPI (Patrick West, Yu Chen) - Ticket mentioned below. Also question of authorization policies -- want users to be able to only see certain information from VIVO. Also are using Drupal for some authentication and creating groups in Drupal to manage this; not so much the authorization piece for editing -- is more about more data or visualizations to display given group membership. Will put together a use case or two or three and will share.
    - https://wiki.duraspace.org/display/VIVO/Creating+a+VIVO+authorization+policy+-+an+example
- Stony Brook (Tammy) - Working on process flow for gathering and transforming information for the new vivo.stonybrookmedicine.edu website, starting with basic demographic information and grants, with publications to come later. Also changed the name from vivo.stonybrook to vivo. stonybrookmedicine to reflect institutional preference and engagement.
- Weill Cornell (Paul) – Going live on January 7. Interested in hearing back about performance testing. Have a new, better-performing server set up, but still concerned about performance under multiple simultaneous editing sessions. Using a server-based performance monitoring tool
    - Ted -- the sysadmin at Brown did testing with JMeter, and developed a suite of tests that included logging in and making edits to a publication; JMeter has lots of tools for simulating 10 users at a time. At Brown the effort was to get a baseline measure of performance.

Notable list traffic

See the vivo-dev-all archive and vivo-imp-issues archive for complete email threads

Trying to subclass VitroHttpServlet to create customized view controller... which file should we change in order to register the new view controller? Also, how do I wire the new view controller that I created to other view controller? And how should I process the VitroRequest instance to redirect to another view controller? (Yu@RPI). If the objective is to go back to specific URL, there are methods in the Generator class that allow specifying where the user should be directed following a custom form (if this is what you are trying to do). Also want to customize the page to edit a property of an instance; there's a post() request from the page but is trying to find the corresponding doPost() method that resolves the post() request sent from the page. (Huda) -- we use a Generator class to accomplish this, normally. (Yu) -- want to change to a multipart post request so can upload images (documents, datasets, whatever) as well as submit data in the response.

For the RPI question, can we get a contact email for both people who answered the two possibilities? The tomcat controller, and the method in the Generator class? hjk54@cornell.edu

Is it okay to have 2 different VIVO instances running on the same server? Can both of the web site use the same solr index or what is the best way to do it? (Gawri@Queensland University of Technology)

Jon: my instinct is one VIVO per Tomcat for anything that is going to production -- we've run multiple VIVOs on development machines but notice that performance degrades when we add significant amounts of data. The general approach if you have to run more than one VIVO on the same server is to have an individual Tomcat per VIVO, running on different ports. For example, Stony Brook runs 2 VIVOs on 2 separate virtual machines.

Call-in Information

Date: Every Thursday, no end date

Time: 1:00 pm, Eastern Daylight Time (New York, GMT-04:00)

Meeting Number: 641 825 891

To join the online meeting

Go to https://cornell.webex.com/cornell/e.php?AT=WMI&EventID=167096322&RT=MiM2

If requested, enter your name and email address.

Click "Join".

To view in other time zones or languages, please click the link: https://cornell.webex.com/cornell/globalcallin.php?serviceType=MC&ED=167096322&tollFree=1

If those links don't work, please visit the Cornell meeting page and look for a VIVO meeting.

To join the audio conference only

To receive a call back, provide your phone number when you join the meeting, or call the number below and enter the access code.

Call-in toll-free number (US/Canada): 1-855-244-8681

Call-in toll number (US/Canada): 1-650-479-3207

Global call-in numbers: https://cornell.webex.com/cornelluniversity/globalcallin.php?serviceType=MC&ED=161711167&tollFree=1

Toll-free dialing restrictions: http://www.webex.com/pdf/tollfree_restrictions.pdf

Access code:645 873 290