

Implementation and Development Call 20130228

Calls are held every Thursday at 1 pm eastern time (GMT-5) – convert to your time at <http://www.thetimezoneconverter.com>
These calls now use WebEx and have no limit on the number of attendees – see the "Call-in Information" at the bottom of this page.

Updates

- Brown (Ted) – still working on data cleanup from the large load; working on a custom form for presentations and submitted a pull request on GitHub – list view on grants wasn't sorting by date.
 - Stephen – liked the Git post on using Python with the Harvester to write directly into the model – lawlesst.github.com/notebook/pyjniusvivo.html and also see <http://lawlesst.github.com/notebook/vivo-rdfalchemy.html>
- Colorado - working to automate scripts for update so not using the GUI – want to use the CSV ingest via the Harvester. Will share the CSV fetch as soon as has written some test cases
- Cornell - Tim is doing work with maps to show researchers' geographic focus; started with Google Maps which was very easy but uses flash so would not work on mobile devices;
 - we also have new triplets (Huda) and a new baby (Brian C)
- Duke (Richard) - working toward early April/May go-live date for School of Medicine – have the Elements instance fully loaded and are pulling pubs again. Duke has two legacy systems with faculty activities in them – FRED and FDF – loaded them last summer but loaded one last time from those systems, and upgraded the version of Elements. Also brought in a lot more of the ReachNC data from Scival
- Florida (Nicholas) – working on requested Freemarker and CSS changes – UF specific.
- Indiana (VIVO team) – got the full dataset from Florida for testing visualizations; are seeing a Vitro banner when load data – once loaded, will be inserting log statements
 - (Tim) – since changed the database, may not be finding the theme – go to the site information on the Site Admin page and change the selected theme.
- Memorial University (Lisa) - have begun a [Knowledge Mobilization Ontology interest group](#) to work on an ontology to model public engagement and are collecting use cases; anyone who is interested modeling connections between the university and the external community
- Penn (John Mark) - have had an internal rollout over the past few weeks and are hoping to go public sometime after mid-March, following one more message alerting faculty. How have other sites configured robots.txt to avoid problems in trying to harvest the site too fast?
 - Brian – could use the IP tables on Linux to throttle or rate limit certain networks – looks like it might be a good option
 - Richard – on another project, a calendar app, register the site with the Google webmaster tools and can turn down the rate at which they crawl; you do have to go back in a few months to keep the rate turned down – so many potential links on a calendar app
 - Brian – changed the 1.5 code so that would not crawl the
 - Ted – is also a crawl frequency directive – John Mark – Microsoft honors but Google may not
- Scripps (Michaelaileen)
 - Problem with two VIVO applications starting when Tomcat starts – Stella is working with Scripps IT staff to resolve.
 - Map of Science issues – It takes over an hour to generate the first Map of Science request for the day but once the first one has been generated, any subsequent ones are generated quickly. Michaelaileen wondered if the problem was due to the recent ingest of 26,000 articles so she ran a test on the UF VIVO that has 36,000 articles and found that Mike Conlon's Map of Science generated in only seconds. At Scripps a cron job runs nightly at midnight that stops then starts Tomcat, so when VIVO restarts as a result, it takes a long time to rebuild the Map of Science cache. Testing the use of a cron job to request a Map of Science after the midnight restart.
 - Penn (John Mark) – Map of Science fails on Penn's VIVO with 200K publications so they turned off Map of Science function
 - Cornell (Jon) – The IU team is working on this issue with Penn
- Stony Brook (Tammy) - revisiting why biosketches are taking up to 6 minutes to run – a result of having a large number of publications. Had been asked to remove the biosketch capability when upgraded to Java 7, and will try to re-activate. With Java 7, are getting an error message in the logs saying there are too many open connections (to MySQL)
- UCSF (Eric) – a question and an observation on the ontology – a lot of people have a preferred name and a given name that may be different; may want to be known as their preferred name but then have name parts for publication disambiguation that may be different. People tend to use their given name for publication – seems like there should be support for this.
 - Stephen – the display is using rdfs:label – but UCSF is using the VIVO ontology with Harvard Profiles, so there needs to be a property in the ontology to use. They do use the label, but that may have additional information such as degrees listed. Makes sense for the UI to have multiple ways of showing the
 - Eric – also need to extend the ontology to support node gadgets
 - Brian Lowe – some kind of custom local extension would be needed for now, but are making a new contact module in the CTSAconnect ISF that will use the vcard contact ontology into nice bundles that can be a set of contact information for a particular context. For example, if you wanted to use a different form of your name with a different position, or to show a different form of your name with certain publications (e.g., the maiden name). Could associate the Vcard instance with the appropriate context.
 - Alex – people were asking about OpenSocial last week in Australia – are there any updates? Eric: the OpenSocial.org website is highlighting the UCSF Profiles now; are waiting for the Apache project to release a new version of the API called Apache Rave as a replacement for Shindig that also supports other ways of plugging in content. Shindig has been very good, and stable, but the developers
 - Jon – Stanford has integrated with Salesforce Chatter – has UCSF? Eric: yes – can follow people and create a group, but are doing it all with OpenSocial. Use OpenSocial to plug in a lot of content from social networking sites – Slideshare, Youtube – and those sites have a network as well, such as Linked In. Thinking about whether there's anything useful to do with that; profiling systems are better at showing really deep data, but modeling current active networks is quite different; more and more people are asking about this.
 - Alex – had a meeting with Atlassian and they are using OpenSocial as a way to integrate content from their own products and perhaps other sites – <http://www.atlassian.com/opensocial/overview>
- Weill Cornell (Paul) - Finished primary ingest of data from individualized Scopus queries, with ten errors but different errors so will be investigating. Will be focusing back on performance. Will be aiming for a public release of VIVO to coincide with the opening of a new research building the end of the year.
- other -

2013 Implementation Fest plans

- Thursday and Friday, April 25-26 have been selected as the dates for the VIVO Implementation Fest.
- There is no registration fee for the workshop.

- This is quite a different event from the August annual conference, but may be used as a way to stimulate more ideas for workshops or presentations
- A community hands-on development day is planned as an optional additional activity on Saturday the 27th.
- The [2013 VIVO Implementation Fest](#) page has basic information on transportation and hotels and will be fleshed out in the coming weeks as the schedule is confirmed
- We hope to have an email out about this in the next few days
- Please send feedback in the next week or so
 - Paul – is there a way for people to attend virtually?
 - Alex – recorded some sessions last year, but not for real-time streaming; could use WebEx and try to get a wireless mike and have someone monitor the chat window.
- The [2012 VIVO Implementation Fest](#) page has additional information about last year's schedule and presentations.

Report from Melbourne

- Event info <http://blogs.unimelb.edu.au/vivoands/2012/12/21/research-profiles-conference-2013/>
- Melbourne Find an Expert <http://findanexpert.unimelb.edu.au>
- Griffith Research Hub – just won an award from Stanford University as an innovative library application <http://research-hub.griffith.edu.au/>
- University of Western Australia – <https://communities.research.uwa.edu.au/vivo/>

Notable implementation and development list traffic

- (Ted at Brown) Custom template for a top-level organization – a workaround at the moment.
- (Jim at Cornell) "At the end of March, I plan to change the VIVO build script and documentation to require Java 7. If you are working from the develop branch, make sure you are compiling and running with Java 7, or you will be unhappy. Let me know if you foresee any problems with this change." More info at:
 - <http://openj.dk/2013/02/21/end-of-public-updates-for-oracle-java-se-6/>
 - <http://radar.oreilly.com/2011/09/java7-features.html>
- (Giuseppe @ University of London) Ingesting research areas from an XML file
 - If I harvest just two people with a common research area, all works properly, but if I harvest a larger amount, the research area is displayed within the person's page, but the list of people is not present on the area's own page – Does this suggest a memory issue (i.e. the inference engine having to calculate inverse properties over a dataset that is too big)?
 - Note that I've made it so the research area is only ever shared by at most two people, so the problem doesn't seem to lay in the inference on a subset, rather on the size of the full dataset (over 40,000 records). Any ideas?
- (Michael @ Colorado LASP) Trying to exactly replicate a complete system from machine to machine; a SQL dump would probably be a good way to handle this on an occasional basis.
 - As for the longer answer to what our actual end goal is... In our shop, we'd like to carry an exhaustive instance of VIVO internally, with both the data we'd like to display publically as well as some sensitive metadata about configuration management, storage hardware, billing information, etc that cannot be exposed. All updates to our metadata information will feed into this instance. Additionally, we'd like to push a subset of this information (that doesn't include the sensitive stuff) into a public space for read-only access, probably on something like a nightly basis. Do you think it be feasible (and scalable) to do a full SQL dump of our DB on a nightly basis to accomplish this? Assuming it were, we could simply store all the sensitive information in a separate JENA model and drop that before we expose it to the public, yes?
- (John @ Cornell) I want to use VIVO for authentication but just use the Persons [vivo:primaryEmail](#) property as the "user name" (externalAuthId) for user accounts ...

Call-in Information

Topic: VIVO weekly call

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](#)

[last meeting](#) | [next meeting](#)