

Implementation and Development Call 20130110

Calls are held every Thursday at 1 pm eastern daylight time (GMT-5) – convert to your time at <http://www.thetimezoneconverter.com>

These calls now use WebEx – see the "Call-in Information" at the bottom of this page.

Please add additional agenda items or updates --

Updates

- Stony Brook – Erich has been working on the WebID authentication for another app and most of the code should be portable over to VIVO, although was done with Jetty instead of Tomcat. Can override the trust manager in JDK1.7 and implement authentication.
 - Alex – might be a good topic for a focused dev call
- Florida –
- Weill - 1) Made some progress to the performance work. We noticed a lot of temp db tables being created during queries. The tool (New Relic) that we use doesn't show the exact queries that are run during the page loads. Anyone know of tools that can display the queries live during page loads? Presumably SPARQL queries get converted to SQL queries during page loads? *Paul – Eliza is working on optimizing queries. Brian Lowe – you can log all the queries at the MySQL level, which can be pretty massive, but is the straightforward 2) Thanks so much for the feedback on db query cache size from last week's meeting.*
 - Paul – pushing to meet a deadline to provide publication data to a separate home-grown faculty evaluation tool, where they will disavow or allow any publications from 2012; using a separate tool to collect feedback to avoid corrupting data from authoritative sources, but will feed change information back to VIVO. Not yet opening up editing, and concerned about multiple entries of the same publications by co-authors – easier to figure out everything on their behalf and then ask the faculty to correct our errors, and have some people with 700+ publications. Also means that even if faculty are not interested in editing, we can provide good profiles.
 - Have revisited Scopus Ids for all faculty.
- UCSF – *Eric is still doing the conversion of profiles and expect to take another few weeks, and then wants to work with a new RDF to JSON standard that has come out in the past few weeks. Had been using the Epimorphics one but have found another – <http://json-ld.org> and <https://github.com/tristan/jsonld-java>*
 - *Converting existing OpenSocial gadgets to RDF and are finding them actually simpler.*
- Duke --
- SUNY Buffalo – Mark – acquired a new server and hoping to apply it to the production VIVO soon
- Scripps – Michael – have done a huge ingest of 26,000 publications from an in-house EndNote library via contract with Stella; have been troubleshooting and correcting some issues in their VIVO installation and configuration
 - Question: How do you handle author disambiguation?
 - A: Scripps has only 298 faculty profiles so Stella added the names to the program which then runs a match against a faculty name field in the Endnote records; as the program finds name variants that have not previously been recognized, they are added to the list of matches to check during ingest; problems commonly arise from suffixes (e.g., Jr., III) or special characters (e.g., publishers vary in how they handle umlauts, Muller vs. Mueller vs. Mu(umlaut)ller)
- Cornell – Tim – adding mini-reports as embedded sparql queries within individual department pages to show graduate field affiliations and research areas
 - Huda – writing up usability page results; and is using the within-page sparql pages; starting on implementing additional search facets tomorrow
- Colorado – Stephen – creating new RDF for the Colorado Springs campus every other day; put out a new version for the Boulder campus VIVO with the new version of Solr that fixes a Solr bug with nicknames (see below).
 - We have talked about creating a VIVO repository of testbed data and would like to pursue that soon
- Penn
- NYU
- Johns Hopkins
- Indiana
- Brown
- other

Other updates or topics

- Katy Börner's [Information Visualization MOOC](#) starts January 22 and is now open for registration
- Implementation Fest plans beginning to take shape
 - Policy track – how does VIVO with with other tools supporting faculty reporting, institutional reporting, campus web strategies, data and publication repositories, and other significant elements of the research university ecosystem?
 - Technical track – community building: working together to fill gaps in the VIVO community infrastructure, address documentation and configuration issues with Solr and Fuseki, and coordinate data ingest and data maintenance strategies
 - Should this be an extra developer day to do some new development, address certain bugs, performance improvement, data ingest, or whatever
 - In person or via web conference? Sentiment on yesterday's ontology call stressed the value of meeting in person, and UC Boulder is graciously willing to host again this year. Should we also make an effort to set up high quality video conferencing for some sessions, or would Webex be adequate?
 - Stony Brook can host a 40-way HD video that supports SIP or H.323 video
 - we should probably have a test
 - How many people think they might be able to attend a 2-day event in mid to late April? (timing not final, but sooner rather than later)
 - Mark and perhaps a couple others from Buffalo
 - Paul can attend from Weill
 - add your name and school here
- From Jim Blake: We are just starting to use the JIRA instance at DuraSpace <https://jira.duraspace.org/browse/VIVO>
 - You can already read the issues. Anyone can read the issues.
 - If you want to create issues, just create an account and log in. (click "Log In" in the upper right, and then click "Sign Up")
 - If you want to edit issues, assign issues, or have issues assigned to you, sign up, send Jim your username, and he will add you to the project.

- Other announcements?
 - <http://vivo.vivoweb.org> is back online again, thanks to Brian Caruso
 - Jim and Alex will be attending the [Research Profiles Conference 2013](#) in Australia February 18
 - may schedule a focused developer call for 3 or 4 pm ET on Thursday, February 21 from the University of Melbourne that could include the Australian development community – a first global VIVO development call.

Notable development list traffic

- As documented in this [mailing list thread](#), UF's problem with re-indexing it's data in 1.5.1 hasn't been resolved. Possible root cause is bad data in MySQL evidenced by repeated errors in logs, but the logs just point to what should be allowed escape characters (linefeeds) that should not cause Java, Jena or Solr to throw exceptions. Should this be the topic of dedicated development meeting? Who wants to be involved and what time slot would work – does the 2 pm ET Thursday slot work for people?
 - Chin Hua also needs to have this bug resolved to be able to debug the visualization issues Florida and Penn have been experiencing
- Problems with non-Roman characters – make sure your server.xml contains **URIEncoding="UTF-8"** in the active <Connector> element. [more](#)
- Transport error in connecting to Solr (resolved – was an outdated configuration)
- Updating Solr to be able to use new modules/add-ons
 - just dropped in the new Solr WAR file and changed the build.config to point to the new version
 - everything worked in VIVO
 - when pushed the update to the Colorado Springs campus, a data update failed so had to wipe the webapps directory for VIVO and Solr, not just the Solr directory – was just giving an internal system error during the add process
 - fixed the issue of searching with wildcards and will be other new features to take advantage of – there's a changelog for Solr online
- Connecting VIVO to a REST-style web service to migrate existing profiles to VIVO
- Generating a list of keywords from VIVO
- Last data update date for the entire site at CU Boulder (are not allowing individual editing, and update the entire site every other day)
 - might in the future want to have a tooltip that would show where more granular data elements in different graphs or even on a per individual basis
 - is this a feature others would like – a more robust version of this might change a number of files
 - Jim – others might or might not need that particular feature, but would love to see something more generalized to make arbitrary information available to Freemarker templates so can put that information on any page, or in the footer
 - we already have an investment in our data-getter class that allows you to configure it with a sparql query and the name of the Freemarker variable you would like to put the returned value(s) into
 - just need to figure out how to have a query run to appear on any page, and how to set up one or more data getters to run every time any page is generated, as opposed to running queries within specific page templates that the controller invokes just for that template
 - is this as generalizable as a key-value pair
 - are using different named graphs as of 1.5.1, so can replace one data element and can diff the DAT files and only re-run ingest processes where there have been changes. Has enabled reductions of the ingest load from 3-4 hours to about an hour
 - can add statements to the graph for provenance on batch updates of the entire graph
- Other issues?

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=16711167&tollFree=1>
- Toll-free dialing restrictions: http://www.webex.com/pdf/tollfree_restrictions.pdf

Access [code:645 873 290](#)

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