

Development Call 20120816

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

Please add additional agenda items or updates --

Updates

- Weill Cornell – 1) Working to customize Browse by, search and index page ... can part of this be done using Internal class? Which is the file that determines what to show in the Browse by section?
- Stony Brook
- Scripps
- Penn
- NYU
- Johns Hopkins
- Florida
- Duke
- Cornell
- Colorado
- Brown
- any others on the call

Followup from GitHub discussion on the last 2 calls

Jim has created a [Transition to new Community tools?](#) wiki page for questions and discussion to date.

Conference Previews

[Schedule at a Glance](#)

I'd like to walk through the program and have anyone on the call who's involved in any workshop, presentation, panel, or poster to very briefly describe the main topic(s) – there are too many sessions for any one of us to attend, and it will be helpful to have more than the titles when making our decisions on.

Notable Development List Traffic

Integrating VIVO with Sakai OAE

(Simon G.)

I've been doing some further work on the Sakai-VIVO integration and have some reasonable success – am able to output full VIVO profiles in json with the freemarker templates and display those on our profile pages, and have integrated VIVO's solr engine with our cluster so we can search vivo data from within Sakai. I am however having some issues with linking vivo profiles and Sakai users. This is probably due to our use-cases and how to properly link datasets but here goes:

I need to know the URI of a VIVO profile that corresponds to a user in Sakai so we can retrieve a profile by that URI. Use-cases:

- A pre-existing fully populated VIVO instance -----> What would be the best way to get the URI of a profile that matches the Sakai user?
- A fresh VIVO instance with no data in it -----> Create a profile in VIVO automatically and get the URI back in the request (this should be doable)

Upon user creation we have the following info: 'first name', 'last name', 'email' and in case we're importing from LDAP, the ldap cn. (I was sort of under the impression that VIVO supported ldap authentication/import but that's not the case I think?)

A side-question I'd like to pose to people who are running in production, how much access is typically given to users to edit their profile? Are they allowed to edit it themselves or does an admin user/harvester take care of the updates? The reason I ask this question is that we're thinking of not letting a user edit his profile which would allow us to cache (almost) everything. That would dramatically improve the speed of getting a profile (~700ms vs 18ms for an average profile.) *If you have sources for the data that needs updating, blocking editing to be able to use caching is a good strategy – we think Melbourne's Find an Expert uses caching, as well as having other interesting customizations*

See also [NIHVIVO-3925](#) for patches Simon worked out for compatibility with Solr 4.0, that does not seem to allow multiple colons in a query string

Glacially slow indexing

(John Mark)

We're running VIVO 1.5 in test on a virtual server. Recently we loaded a large batch of data on publications, many with abstracts, from PubMed and an internal database, using the RDF upload feature. The publication file was quite large (a 667 MB N3 file for ~200k publications; since we'd originally generated this before we installed 1.5, it includes precomputed inverse statements.) Despite the size, the data went in relatively quickly. (I checked back a few hours after I uploaded the file, and could find data entities from both the start and end of the file.) We gave the Tomcat Java processes a good amount of memory, which seems to have helped.

However, once we started building an index of this data, performance suffered badly. The indexer reported "Number of individuals to be indexed: 548159 by 10 worker threads", but then the time per individual started at 22900 msec and went up from there. Our virtual server admin reported the server was hitting the SAN disks very hard (enough to disrupt other virtual servers). When we finally shut down things down entirely, it had indexed 32300 individuals at a rate of 36512 msec per individual, with a load average of over 14. Has anyone else seen behavior like this, and if so, what can be done to fix it? (Indexing the smaller datasets we loaded earlier went much quicker.)

Other issues

- 1.5 page management (Tammy and Tim)
- preventing CV generation via robots.txt – if you are using apache for either AJP or ProxyPass you need to place the entire robots.txt into the document root of your default site (Vince)

Items for next week

Developer Community Tools and Web Presence BOF Thursday at noon

There will be no call next week due to the VIVO Conference, but Jim Blake will be leading a Birds of a Feather session from noon to 1 pm on Thursday where some of the recent conversations about transitioning the wiki and moving our code repository from SVN to Git (and from SourceForge to GitHub) and any other issues you want to bring up can be discussed in person. The location will be in the program – it will be in one of the session meeting rooms.

Call-in Information

1. Please join my meeting. <https://www1.gotomeeting.com/join/322087560>
2. Use your microphone and speakers (VoIP) - a headset is recommended. Or, call in using your telephone.

Dial +1 (773) 897-3008
Access Code: 322-087-560
Audio PIN: Shown after joining the meeting

Meeting ID: 322-087-560

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