

# Development Call 20121206

Calls are held every Thursday at 1 pm eastern daylight time (GMT-5) – convert 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 --

## Rethinking VIVO calls

*Discussion during the meeting:*

- Keep this call but describe it as what it is – a blend of implementation and development; reduce overlap with implementation call so that people focused on implementation don't have to attend both
- Arrange focused calls on topics that might be about either implementation or development, or that bring in experts from related semantic web projects such as Fuseki
- It may be hard for people to justify attending more than one call per week, so be clear on what the topic of each supplementary call would be

## Suggested special topics so far

Effective use of Git

- what's a good workflow as a core developer, and how do you fully integrate with the Git model?
- if on the implementation side, how to contribute
- are we going to use GitFlow?

Extending search functionality

- following on the [Griffith Research Hub](#) example of adding additional facets to search results
- building a local version of [as for example when running two instances of VIVO for two campuses of your university, or to link in outside institutions such as the national labs around Boulder for a Colorado research hub

OpenSocial and <http://www.orng.info> Open Researcher Networking Gadgets.]

Using the three-tiered build (Vitro + VIVO + your local modifications)

- Colorado's implementation and further work they have done using Git for deployment to test and production servers
- Consideration of whether a three-tiered build would make sense for the ontology

## Background prior to the call

Part of the discussion on last week's implementation call concerned making a clearer distinction between calls focused on people's activities and experiences with populating and configuring VIVO vs. calls more focused discussion of code development and/or debugging or modifying an enterprise VIVO installation.

These weekly calls have been a mix of both types but with a more frequent emphasis on the former – and we think that as we operate as a community-driven project than a grant-directed project, developers would like to see more of the latter. Is this true, and if so, how should we best accomplish this?

One option would be to use the weekly Thursday 1 pm time slot for updates on new development, data ingest procedures, semantic web tools, implementation experiences including demos, and suggestions for improvement – a combination of the current development and implementation calls, but given a different name so that non-programmers feel comfortable attending. Alex and I are willing to explore this new model, but we need to hear from you.

This opens up the option of holding separate, likely biweekly calls on one or two single topics in more depth, since our updates now often take up at least half the hour. Announcing topics with more advanced notice might encourage attendance by those wanting in-depth technical information and having less interested in general updates.

One way to test the waters for these new focused calls would be to call for topics and assess interest based on whether anyone steps forward to organize a half hour or hour on each topic – not necessarily as the presenter. Jim Blake has done a number of successful presentations that blend code, explanatory slides, and demos, but other patterns would work, including researching a topic such as WebID or recruiting someone from outside to speak and answer questions, as Alex has done.

The ontology calls will also continue as they address a different need, as do the outreach and adoption calls that Kristi Holmes leads.

If you can't attend tomorrow's call and have opinions to share, feel free to send a note to the list or write Alex Viglio or me directly.

## Updates

- Weill – Paul - working on quality for publications, and on performance – checking
- UCSF – Eric – still working to upgrade to the RDF version of Profiles. Will be looking for a more standard way of converting RDF to JSON in the next improvements to OpenSocial in VIVO. Knows the [Plumage](#) people and one of the targets for next year is to better align the UCSF Profiles with core facilities
- Penn – John Mark – load test prior to internal rollout with simulating lots of simultaneous edits, and one of the Java processing pegged and didn't go down even when the load went away. Looking at separating database from tomcat. Anybody using MemCache or Squid? The visualizations

*seem to be the problem – have an author with 900 publications, and even without server load get a service unavailable measure. His profile page comes up fine."*

- NYU
- Johns Hopkins
- Indiana
- Florida - Nicholas – Got the URI tool closer to working, but still some issues – enters an EISSN but doesn't show up – is there an order that the additions and retracts should happen in?
- Duke - Richard – working on speeding up ingest of publications from Symplectic
- Cornell – Tim is preparing 'short views' with more detailed, list-view type information in search results and index pages
- Colorado – doing a soft launch of VIVO for the Colorado Springs campus behind the firewall, and hoping to roll out 1.5.1 next week. Have the 3-tier build working and likes that way of building. Will work with Jim on a wiki article. With managed services team blessing, can push releases out to dev and eventually production server from developer workstations – can roll out changes via a standard process that's the same for both a test server and for production, all using Git and post-receive hooks for Git (a bash script, basically). Also useful for supporting two campuses.
- Brown – Ted – Released VIVO instance to faculty on Monday and they have a period of time to look at their profiles – one major problem is with presentations. If try to edit an existing presentation from the profile page, it hangs and the server becomes completely maxed. Working on isolating the problem – is this something with initializing the autocomplete? Turned the logging up on the development server – a large sparql query was the last thing that was running; looked at the sparql queries involved and they did not exhibit the same behavior (the custom form works on a smaller subset of data).
- other

## Many dimensions of performance

As a VIVO transitions from test into production it's not uncommon to encounter performance issues that may be attributable to system configuration, differences in server memory/processing power/OS, search engine traffic and robots.txt settings, amount of data loaded (including data not successfully removed), features enabled or not enabled (especially visualizations), and outright bugs. What do we collectively know about these factors, and how can we help each other identify, address, and document them?

We also recognize caching is a closely-related issue, and Arve Solland gave a [talk](#) on Griffith University's approach to caching at the 2012 VIVO Conference. When profiles get very large, the database reads required to assemble the data for a VIVO page may be too slow, even if the software were perfectly efficient, which it is not. Strategies that cache the HTML generated for a page offer the most promise for near-constant-time (fast) page loads, but the cache of a page needs to be expired when any edit in VIVO will result in a change to that page. Think of a star professor with 350 publications averaging 5 authors each – if any author adds a middle initial, the star professor's cached page should be expired.

Each "document" in the Solr index (corresponding to any individual page in VIVO) includes a field populated with date and time of the most recent change to trigger a re-index of that individual, even if that change was not to a statement with that individual as its subject – as would be the case with names of co-authors. In theory this Solr field, if made available in standard HTTP headers, could indicate to a caching tool such as [Squid](#) or [memcached](#) when a page has expired and needs to be re-generated as opposed to having Apache render the page from the cache.

There's enough meat here to discuss that we may want to address over several calls – suggestions welcome.

## Wiki move update

Jim Blake will be migrating this wiki over the coming weekend, and will disable editing late Friday, December 7. The plan is to have the same content visible and looking much the same on the DuraSpace Confluence-based wiki as of Monday, December 10.

*Anything contributed to the wiki after Friday close of business will not be migrated – Jim will attempt to disable editing to avoid losing any content.*

*If you use the same account name on the DuraSpace wiki as you did on SourceForge, you will be seen as the owner of any page you created, but this may not be important since anyone can edit.*

*Does not look as though history can be moved over. Conversion will not be perfect – things like tags inside code excerpts may not come over correctly – look at your favorite pages after conversion to confirm.*

## Notable development list traffic

- removing "Person" from the browse list on the home page (not the People menu page, where the Person class can be unchecked via page management – requires changes to the browseClassGroups.js file \_(Weill, Stony Brook)\_
- upgrade halted in FileGraphSetup (Northeastern) – Jim suggests rebuilding the search index, which when there's no content in the system should only take a couple seconds
- optionally creating role in EditConfigurationGenerator – Melbourne is doing advanced custom forms development
- solution to missing inverse property (Brown) – They introduced a local sub property of bibo:performer that has the desired inverse to link the performances from people pages
- VIVO-Fuseki-Elda – making Fuseki accept sparql updates, and working with VIVO, Fuseki, and Elda [Epimorphics Linked Data implementation](#). Michael from the LASP lab at Boulder will keep working the the latest Fuseki code and see if anyone in that community can explain its failure to update, and bring back anything he learns.
- WebID in VIVO? (Stony Brook) – anyone interested in WebID should contact Erich Bremer
- Dpedia countries showing up in search results – Alex suggest the option of a three-tiered or more modular build approach for the ontology, too, which fits with the modularity that the [CTSAconnect](#) project is developing
- VIVO and CAS authentication (Stony Brook and Notre Dame) – Tammy reports this is working
- what's (still) up with <http://vivo.vivoweb.org>? – still in transit, somewhat delayed by other deadlines; primarily just need to transfer the domain name
- anything else?
  - Joe Mancino would like to hear from others about how they are doing load testing, especially for simultaneous editing. Jim has done load testing in fairly simplistic ways as part of our release preparations, but considers it still rudimentary.

## Still pending

- map of science and temporal graph visualizations – *Chin Hua from the visualization team at Indiana University wrote to say he is working on another project but will get back to improving the visualization caching work in December*

## Call-in Information

Topic: VIVO weekly developer call

Date: Every Thursday, no end date

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

Meeting Number: 645 873 290

### To join the online meeting

- Go to <https://cornelluniversity.webex.com/cornelluniversity/j.php?ED=161711167&UID=1278154642&PW=NWWM4ZDNhYWZm&RT=MIMxMQ%3D%3D>
- If requested, enter your name and email address.
- Click "Join".

To view in other time zones or languages, please click the link: <https://cornelluniversity.webex.com/cornelluniversity/j.php?ED=161711167&UID=1278154642&PW=NWWM4ZDNhYWZm&ORT=MIMxMQ%3D%3D>

### 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://cornelluniversity.webex.com/cornelluniversity/globalcallin.php?serviceType=MC&ED=161711167&tollFree=1>
- Toll-free dialing restrictions: [http://www.webex.com/pdf/tollfree\\_restrictions.pdf](http://www.webex.com/pdf/tollfree_restrictions.pdf)

Access code:645 873 290

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