

# Development Call 20120621

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

Please add additional agenda items or updates – we always welcome suggestions

## Updates

- Weill Cornell – 1. Downloaded DVDocs, dvdocs-binaries, dvdocs-documentation, dvdocs-extra-jars from Digital Vita (<http://loafer.dbmi.pitt.edu/dvdownload/dvdocs/>). VIVO plugin - any documentation or tips?
  - Rebuilding the search index takes 20 hours – at NYU internal application (a Rails app) can rebuild just part of the data; easy to rebuild a specific
  - Eliza – would mostly want for publications (have over 30,000 publications, but have added 100 more to it) – would it be possible to send a list of URIs to be indexed?
- North Texas –
- NYU – have a VIVO populated and starting to make a case to university; working with Brian on refining the Solr document to improve the results for faculty to better meet their expectations; for NYU, the default search is pretty agnostic to the type of data, and what they are coding is primarily for people – if put in arthritis, will get the top person at NYU. Focusing more on the semantics of the search than doing a straight boost. Will redefine what a document is for each Person to rank the people if have a ton of publications in one area, for example. Want to focus on the specific use case, rather than indexing events and other less important data. Internally have a RAILS application with all the MeSH terms for all publications from NYU already. Are also very explicit in their search interfaces so people know whether they are searching MeSH or the title and abstract of the publications, rather than interceding to try and determine the correct boosting for different queries – give people 2 boxes and let them try the two approaches – an open hypothesis whether the single search box or separate search boxes work better for people. Also run analytics on clicks in the search results, and are working on a click-through ranking system (based on work by Thorsten Joachims at Cornell) to get more recent work to show up higher. Will check in these ideas back to Subversion.
- Stella – can't find Mike Conlon
- Yin – there is a query parser in Solr to allow that
- John Mark Ockerbloom – the bug with capital letters in the middle of names is still important to fix
- Johns Hopkins –
- Indiana –
- Florida – Getting a few harvests ready for production and upgrading VIVO version
- Duke –
- Cornell –
- Colorado – 1.4.1 update and adding new data in prep for last campus release before public release; checking with UF about Web of Science publications download
- Brown –
- Penn – changed the case of issn property to match better; still working with eissn matching
- others –

## Update on 1.5 release plans

- upcoming code freeze
- meeting Friday to list new Selenium tests needed
- /noscript more common

## Walkthrough – setting up an alternative triple store for VIVO via the new RDF API

Stella Mitchell is testing the new RDF API on an [openrdf.org](http://openrdf.org) Sesame triple store and will walk through the process to set up Sesame and configure VIVO to use it

- VIVO 1.5 will ship with 2 implementations of a new interface for RDF reading and writing
  - Jena SDB as the default (the current triple store used by all VIVOS)
  - openrdf-sesame
- download 2 WAR files from openrdf-sesame – triple store and workbench
- set up a separate tomcat
- follow documentation for Sesame server implementation
- end up with a SPARQL endpoint URL that then have to configure in VIVO
- new deploy.properties setting VitroConnection.DataSource.endpointURI =
- still need to have a Jena SDB model for user accounts, so the deploy.properties relating to Jena remain the same
- can separate writing endpoint from reading to facilitate access control
- log shows which endpoint it's using

## Notable Development List Traffic

- Pending investigation: Recommended search boosting techniques having no effect
- Update on testing at UF of revised SPARQL queries for "rich export" function sending data to CV or biosketch
- Custom form work (Tim Sullivan) – correspondence between the template and N3 optional/required specified in the companion Java generator file

- Bots causing errors that generate emails to the VIVO system notification address
- Specifying names for matching in the Harvester – can be broken out into separate queries, but can handle hundreds. Can also query by affiliation.
- Anybody working with data from grants.gov? Sunita Koul at Washington University has worked with NIH RePorter data – we are not aware of anyone working with grants.gov. Yin from NYU has worked with NIH RePorter for other systems and is willing to help.

## Items for next week

- Report on early testing of version 1.5

## 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](#)