Implementation and Development Call 20130307

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
  - staff meeting with departments regarding internal rollout
  - modeling humanities data that requires some customization (mostly in the SPARQL queries used to generate pages)
  - submitted a pull request
  - also using GitHub for their deployment at https://github.com/Brown-University-Library/vivo ("We are making minor changes to listViewConfig queries and templates")
- Colorado
  - working on conversion from Selenium data load scripts to Harvester for increased automation
  - plan to update develop branch of Harvester with CSV functionality, may roll out release
- Cornell (Jim, Brian, Tim and Jon)
  - Tim working on a map feature (viz using Leaflet JS library http://leafletjs.com/) for the home page – global, U.S., and NY state views
- Duke (Sheri, Richard and Jim)
  - targeting April date for School of Medicine
- Florida (Matt and Nicholas)
- Indiana
  - Investigating issues with visualization Map of Science, e.g. out of bounds memory exception – has anyone worked on the Map of Science that has data or notes?
- John Hopkins University (Kelly and Josh)
  - Working on Harvester PubMed ingest (45k records) issues using example harvest – fetch issues until bumped up batch size to 50k, 35 hours for 60k ❓ people, 2nd run (morning after the 1st completed) threw a Java stack overflow error at about 32 hours and only 1/3 complete
  - Stephen suggested commenting out the "remove data" step at start of example script, use transfer to force two TDB models to increase performance, and adding scoring (which uses TDB) to not review old publications
- Memorial University (Lisa)
  - 6 months into 2 year project to do a fit/gap analysis and then how to integrate Yaffle and VIVO concepts at 17-campus university
  - 3rd working group meeting every Friday afternoon (link Jon's wiki page here) on modeling institution's persons and activities along with community person and activities (knowledge mobilization ontology aka public engagement, a 2-way knowledge exchange) also modeling different forms of collaborative activities and outcomes, exploring the lifecycle of an idea (suggested and brokered in Yaffle), identifying knowledge brokers
- NYU (Yin)
- Scripps (Michaeleen)
  - Senior administrators want to know if any VIVOs are being developed to house an institution's entire publication output (primarily journal articles but patents are also of interest). Michaeleen sent a message to the discussion list.
    - Brown started with faculty CVs that include historical data
    - APA and NCAR (eval/demo stage) interested in representing historical publications
  - Map of Science performance: Stella set up a cron job. The following 4 commands have been scheduled in cron (under root) for the past few days and have been working as desired, so that when we request any Map of Science visualization first thing in the morning, it comes up right away.

    - 0 1 * * * /usr/bin/Xvfb :1 &
    - 9 1 * * * DISPLAY=localhost:1 /usr/bin/firefox http://vivo.scripps.edu/vis/map-of-science/CravattBenjamin &
    - 0 3 * * * pkill -f firefox
    - 2 3 * * * pkill -f Xvfb

    The first starts up a program that can function as a display since Firefox requires a display, the second makes the request for a Map of Science visualization to vivo, and the third and fourth shutdown the Xvfb and Firefox processes. There's no error-checking or robustness in this procedure. If Xvfb failed to start up properly one day for some reason, then the process wouldn't work, there's no checking built in here to look out for scenarios such as that. We're going to see how this works before we look at more sophisticated solutions.
  - Co-Author Network visualization performance: We don't think this visualization involves caching in the same way as the Map of Science. Co-Author viz for faculty with a couple hundred citations take one to two minutes to run. Will IU's investigation into Map of Science performance also address performance issues for the Co-Author Network visualizations?
- Stony Brook (Erich and Tammi)
  - Erich attended 6th conf on computational biology http://www.iscb.org/cshals2013 ("CSHALS happens each year about the same time and the same place in Boston"), and speaker question gave Erich an opportunity to pitch VIVO
  - debugging biosketch generation (depending on # of publications per person) – group feedback indicated version of this code from Pittsburgh on GitHub was working with Jena models
- UCSF
  - getting ready for AMIA joint summit conference http://www.amia.org/jointsummits2013
  - launching a YouTube ORNG gadget – will try get it working in VIVO
- WashU (Kristi)
- Weill Cornell
  - troubleshooting performance, meeting to look into ingest of PH, describing positions more accurately, sent out notice to ~500 researchers related to Pubmed data

http://www.discourse.org/

2013 Implementation Fest update

- The VIVO Implementation Fest will be held in at the University of Colorado Boulder on Thursday and Friday, April 25-26
- A community hands-on development day is planned as an optional additional activity on Saturday the 27th, e.g. working on internationalization

- There is no registration fee for the workshop
- **DRAFT** schedule at https://docs.google.com/spreadsheet/pub?key=0AqosQY8NM9BGdEdXWW5yY3FfczdnWC1ZZi04S3E2Smc&output=html
- 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
- The 2012 VIVO Implementation Fest page has additional information about last year's schedule and presentations.
- If you have suggestions for topics, but can't attend the Implementation Fest – do consider developing a workshop for the annual conference in August

2013 VIVO Conference

- The 4th annual VIVO Conference will be held on August 14-16 in St. Louis
- The **Call for Workshop Proposals** should be coming out in the next day or two with a deadline to submit a workshop proposal by **March 29**

DuraSpace VIVO Incubator Update

- Griffith University is now VIVO Sponsor with a 2-year commitment
    - SPIRL award link http://vivoweb.org/blog/2013/03/griffith-universitys-research-hub-receives-spirl-commendation-merit

Notable implementation and development list traffic

- (Tammy @ Stony Brook) digital vita web service - wiki needs to be updated
- Do we want to discuss threads from GitHub (issues, pull request, notifications) here? For example Ted and Jim discussion about grant sorting on profile pages.
- Seemed like a quiet week. Last week's list below.
- (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