Implementation and Development Call 20130516

Updates

Brown (Ted)

- reworking some of our CV ingest related to presentations and service. The data is too variable to assert Events or Committees consistently. Using the label to store more data. Thinking about how this affects self-editing forms and new data
- corresponding with Jim about caching
- doing some matching of local city and state references to FAST (http://www.oclc.org/research/activities/fast.html). Looking promising. Allows us to also link to VIAF and Geonames. Right now working in batch, but might be possible to incorporate into a form.
- also looking at tools from the Semantic Web community, like Silk (https://www.assembla.com/spaces/silk/wiki/Silk_Workbench), that might help manage VIVO data moving forward. Pedro Szekely had referenced Silk at the I-Fest and in some of his papers.

Colorado (Stephen and Alex)

- Completed CSV search results Download Option (submitted pull request)
- Automated Ingest getting a once over and developing new comparison tool to analyze potential defects in the process ahead of release to production
- Symplectic plans moving forward
- Upgrading to a new production server soon and will use the current server for testing new releases, etc.
- SPARQL End Point is ready for release pending new Domain Names
- Creating new Visualizations for VIVO based on research interests and other data in CU Boulder. Will analyze and submit pull requests if any are potentially helpful for VIVO. Anticipates using the Google visualization code base, or d3.js, but may re-use some parts of the existing code to pull data from VIVO (written at Indiana).
- Alex -- Simon has recently expressed interest in sharing the new visualization code from the Melbourne Find an Experts system

Cornell (Jon, Jim, Tim, Chris, Brian L) Tim, Jim, and Brian are working on the 1.6 release -- caching, multiple language support, and adapting the code base for the Integrated Semantic Framework ontology.

Duke (Sheri and Richard)

- http://scholars.duke.edu is now live! Sent the email to faculty on Tuesday after a soft launch Monday -- last minute work to make sure the correct faculty were associated with centers and other details. People have been going through to hide keywords that came over from Symplectic Elements that had been accumulated from keywords on their publications.
  - Are using Google Analytics Real Time -- big spike in "active sessions" with the release
  - Moved users from the staging instance to the production instance via a Tomcat restart. Went through the service review board prior to launch testing the time taken to load pages, and haven't been seeing significant delays since launch.
  - Production environment uses 2 virtual machines with 4 CPUs and 8 Gb memory, running Tomcat with a 4 Gb heap. One server behind an F5 load balancer with admin tools and MySQL; might at some point split the database on its own server. Had at some point added additional memory to MySQL. Thinking of adding another server in the future for a SPARQL endpoint.
- Here is an example of someone embedding their publications into their website. This is a live feed: http://sites.duke.edu/kornbluthlab/publications/ . The "builder" is the tool that you can't see unless you login. We can demo this if you would like.
  - can embed publications, grants, or courses data in other web pages at Duke
  - publications are broken down into separate entries, with separate authorships for Duke authors and a separate author list for others. Publication citations for sharing on other websites via the Duke VIVO widgets are re-assembled from the VIVO data.
  - Richard demoed VIVO Widgets at end of call -- see vivo_widgets package in GitHub https://github.com/OIT-ADS-Web/vivo_widgets as a starting place for sharing this with VIVO community

Memorial (Lisa) No major news this week -- continuing to move ahead.

Scripps (Michaeleen) -- Has authorization to ingest a full five-year history of NIH grants (2008 forward). Will investigate the possibility of modifying the ontology to reflect the many NIH award types that can relate to a single master award from NIH -- would like to be able to show all the sub-awards and related awards connected to the primary one.

Stony Brook (Tammy) -- more updates to http://reach.suny.edu and moving forward with VIVO for the full Stony Brook campus, including identifying new data sources and some local ontology extensions. And still moving forward to create the WebId implementation.UF (Nicholas) -- as part of the CTSAconnect project, was able to load the clinical module into a VIVO and was able to query against that to generate tag clouds and other visualizations that Chris Barnes wrote using the Google visualization API.

- the first (beta) release of the Integrated Semantic Framework ontology is slated for Monday -- stay tuned

Weill Cornell (Eliza and Paul) -- Developing workflow to automate cleanup of publications and update citation counts -- the last time the query was run against Scopus took 7 hours to get counts from 40,000 publications. Would like to have an update option through the Harvester rather than first deleting and then re-ingesting.

- Also working on figuring out person types -- have specific definitions of full-time faculty, voluntary faculty, courtesy faculty, instructors, postdoctoral associates, research associates etc. that are trying to match to the actual available data source. Not currently working on the notion of faculty "track"

Upcoming Events

VIVO in New York City May 20

- 2-4 pm at 1305 York Avenue in the Greenberg Center – about 10 institutions have RSVP'd so far

VIVO Webinars (all Tuesdays at 11)

- The "Overview of VIVO" webinar was held May 14: presentation slides webinar recording
- June 4 -- Case Studies: VIVO at Colorado, Duke, Brown and Weill Cornell Medical College. **Register HERE**
- June 11 -- VIVO Technical Deep Dive. **Register HERE**

Notable implementation and development list issues

- Fuseki behind Apache (Stephen)

    - Requires a unique domain name example -- vivo.colorado.edu/fuseki(broken) fuseki.vivo.colorado.edu(works).
    - possible opportunity for our community to contribute to Fuseki project, e.g. use case/requirements, documentation, extend code to support /fuseki behind Apache, extend code to fix limited "pages" implementation (makes several assumptions, .tpl mime type issue)
- Scholars @ Duke – can we see what sharing publications to another website offers to a logged-in user? uses the code development on the VIVO Widgets - grant project mini-grant in 2011

    - uses upstart to run in a separate JVM
    - code is on GitHub -- VIVO widgets -- will need some improvements to the listener to real-time update of its separate Solr index
    - people want overviews, contact information, keywords -- people like the JSON format
    - can in the future consider a SPARQL endpoint for more flexibility
- Making a downloadable CSV of search results (Stephen) - Pull Request
- see also VIVO 1.6 release planning in this wiki

    - multi-language and caching support are significant aspects of the work

Call-in Information

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