2015-09-25 Steering Group Minutes

September 25, 2015, 1 PM EST

Attendees

Steering Group Members

Paul Albert, Jon Corson-Rikert ⭐, Melissa Haendel, Dean B. Krafft, Robert H. McDonald, Andi Ogier, Bart Ragon, Alex Viggio

⭐= note taker

Ex officio

Jonathan Markow, Mike Conlon, debra hanken kurtz, Graham Triggs

Regrets

Kristi Holmes, Eric Meeks , Julia Trimmer, Dean B. Krafft

Dial-In Number: 641-715-3650, Participant code: 117433#

Agenda

	Item	Time	Facilitator	Notes
1	Updates	5 min	All	
2	Review agenda	2 min	All	Revise, reorder if needed
3	Welcome Graham	15 min	Mike, All	Graham Triggs joined the project as Technical lead on Monday.
5	Semantic Versioning	15 min	Mike, All	Would like to use semantic versioning for VIVO. See https://goo.gl/i04Z02. See http://semver.org/.
4	Some thoughts on a next release	15 min	Mike, All	Discussion of next release
7	Future topics	5 min	All	attribution/contribution efforts (10/16); how does VIVO get bigger?; training program; rotation of Steering Group members

Notes

1. Updates
   a. Upcoming meetings – CNI, ISWC, NDC
      i. 4th National Data Service Consortium Workshop, October 19-21, San Diego Supercomputer Center, http://www.nationaldataservice.org/get_involved/events/NDS4/
      i. will anybody be attending the National Data Service meeting October 19-21?
      ii. Please let Mike know if you are attending any meetings where VIVO will be presented and/or by topic are relevant to VIVO
   b. Justin Littman's new service
      i. very much along the lines discussed for different APIs in the roadmap discussions (Justin is a member of that task force)
      ii. See https://github.com/gwu-libraries/vivo2notld
   c. Chris Barnes, new data
   d. New Relationship diagram – teaching
   e. Community pages facelift and blog post
2. Review agenda
3. Welcome Graham
   a. Joined DuraSpace on Monday, and he and Mike have had extensive conversations
   b. Has already been doing some interesting work
   c. Graham: joined from Symplectic, having done repository integration there as well as rewriting and updating the connector to VIVO-ISF via the Harvester; 20 years working in online publishing
   d. Introductions all around
4. Semantic Versioning for VIVO
   a. VIVO does a 3-part version number – e.g., 1.8.0
   b. Ran across in exploring GitHub – the founder wrote down his thoughts on major versions, minor versions, and patch versions – and went to the trouble of defining what this numbering system might mean. See http://semver.org
   c. Having a guideline for deciding something is a major version, minor version, or dot (patch) version would be helpful for VIVO
   d. Provides explicit information about backward compatibility and the existence of new features, but there's an element of marketing that impacts releases, too
      i. Once worked on a project that never got past their 2.x.x level, and after a while was perceived as being stagnant since never got to 3
      ii. Make allowances for when you need to give a burst of energy and step forward

e. From a marketing point of view, VIVO may have an opposite problem – it's often the case that there were ontology changes in almost every version that did provide
   i. The delivered software for upgrading the database and converting to new triples has always been strong
   ii. But there are changes to other things that relate to VIVO, and warranted major version numbers to warn people that the relationship of the app to other systems has changed
      1. Consequently our major version numbers would go up rapidly without demonstrating new features
      2. We may decide to focus on alerting people to schema changes
   iii. And no community can sustain a lot of disruptive changes that take a lot of work very frequently
      1. When the Fedora community did their major architectural rewrite, the goal was to position it as something for new users rather than as an update – and they spent the next year doing migration support and building scripts to help people with old scripts get up to the new version, including prototypes with pilots;
         a. they socialized that process into the community
f. We have versioning of both the ontology and the application
g. The ontology sits outside the application, and we should track the provenance and versioning of any component
   i. The ontology group is getting some core developers involved, but to move forward the ontology and tools for working with the ontology
   ii. But it's not working on consumption of the ontology by the application – should be a separate task force, perhaps led by Graham
h. If these are separate tracks, how do you ensure that the ontology and software don't develop incompatibilities?
   i. A standard problem – ideally you have somebody who's knowledgeable about the ontology's consumption by the application working with the ontology team
   ii. There's not a very good intermediate layer that buffers the two – other applications have this, but not VIVO
   iii. The VIVO-ISF ontology includes a lot of material not relevant for a lot of VIVO users, but if we continue to be compatible with the larger ontology, we can combine data
i. VIVO is using many different ontologies, and we have a very close working relationship with some of these but not others (e.g., not with VCard)
   i. This buffering process and vision of versioning of the ontologies and the application is important for our planning process
   ii. We are not dependent on the ontology releases necessarily – we can elect not to take certain changes in the ontology into the application
   iii. We can control things that are in our own namespace, but not others; we are constantly
   iv. And we need migration planning and support as part of our release development and support processes
j. There are consequences in SPARQL queries, training, etc. – need to understand and manage them in an appropriate way
   i. We can't issue ontology changes every two months – we need to calibrate the kind of changes we want with our resources and community
   ii. While there are marketing considerations, would like to align those with semantic versioning as far as possible
   iii. We are likely to have all three types of versions, including versions that change the ontology
   iv. Not sure the marketing problems will be encountered in the next several years
k. Mike will put the semantic versioning proposal out to the community for feedback and use cases that we need to consider

5. Some thoughts on a next release
   a. There is a roadmap process that was described at the steering meeting at the conference and described in a poster shared with the community.
   b. There has been some preliminary work around performance problems that might produce a 1.8.1 patch release, based on issues that have in part already been addressed
   c. Mike and Graham went through the open JIRA issues and found another few changes that are more cosmetic or minor improvements, similarly appropriate for a patch release – that would help notify the community that progress is being made
   d. Maven – an idea that Graham has had is that we look at how we deliver the software with an eye toward how we can help automate the creation of a development environment
      i. There are bunch of dependencies that may be able to be done in a more scriptable hat is familiar to Java developers
      ii. Maven is the way developers most commonly support a build process in an integrated development environment
      iii. Would like to make Maven the way to do the Vagrant build process as well, so it's more closely tied to the primary build process – the work to set up the Vagrant instance has to be re-done with each release
         1. Want something that works out of the same repo with the same components and configuration
   e. Other thoughts?
      i. What about benchmarking performance with a standard set of data – page load times; should be real
         1. Ted Lawless has a dataset – we use that to provide some benchmarks; if somebody else downloads and builds it, they can be compared
         2. Chris Barnes has just published data from UF
            a. A single RDF file with 23 million triples that would certainly exercise the application; one feature of the Florida data is that it has things in it that are not expected – both a blessing and a curse – some of the things in the data are very unfair to the application (e.g., that a person is also a journal)
            b. This data might help stimulate analysis and data consistency work as well as performance issues
         3. The most interesting metric is what the delta is between the legacy version and the new version
         4. Run a benchmark on 1.8 and then on 1.8.1, and distribute the results with the release
         5. One goal would be to point out configuration issues in local installations, if there's a wide gulf between what the developers are achieving and what a site installation is achieving
   f. Mike would like to get the roadmap task force together with Graham to talk about the next release – the benchmarking issue is certainly worth discussing
6. Future topics

Action Items

- Mike Conlon will put the Semantic Versioning proposal out to the VIVO community for comment