

2020-07-07 - VIVO Development IG

Date

07 Jul 2020

Call-in Information

Time: 11:00 am, Eastern Time (New York, GMT-04:00)

To join the online meeting:

- Go to: <https://lyrasis.zoom.us/my/vivo1>
- One tap mobile:
 - US: +16699006833,,9358074182# or +19292056099,,9358074182#
- Or Telephone:
 - US: +1 669 900 6833 or +1 929 205 6099 or 877 853 5257
 - Meeting ID: 935 807 4182
- International numbers available: <https://zoom.us/j/aeANHanzED>

Slack

- <https://vivo-project.slack.com>
 - Self-register at: <http://bit.ly/vivo-slack>

Attendees

★ Indicating note-taker

1. [Benjamin Gross](#)
2. [Nicolas Dickner](#)
3. [Michel Héon](#)
4. [Huda Khan](#) ★
5. [Don Elsborg](#)
6. [Rachid Belkouch](#)
7. [William Welling](#)

Agenda

1. Welcome to the VIVO Committers team: [William Welling](#)
2. [i18n Sprint updates](#)
 - a. Mini-sprint focused on VIVO committer review/refactor: July 14-16th
3. Moving priorities forward: Data ingest
 - a. What are the use cases?
 - b. What the "entities" to be ingested? ([reference](#)) ...let's start simple
4. [2020-07-15 - Special Topic - VIVO Scholar Next Steps](#)
5. ...

Future topics

1. [Vitro JMS messaging approaches](#) - redux
 - a. Which architectural pattern should we take?
 - b. What should the body of the messages be?
2. Incremental
 - a.

Unable to locate Jira server for this macro. It may be due to Application Link configuration.
 - b.

Unable to locate Jira server for this macro. It may be due to Application Link configuration.
 - c.

Unable to locate Jira server for this macro. It may be due to Application Link configuration.
 - i.

Unable to locate Jira server for this macro. It may be due to Application Link configuration.

Tickets

1. Status of In-Review tickets

type	key	summary	assignee	reporter	priority	status	resolution	created	updated	due
------	-----	---------	----------	----------	----------	--------	------------	---------	---------	-----

Unable to locate Jira server for this macro. It may be due to Application Link configuration.

Notes

Draft notes in Google-Doc

1. Welcome to the VIVO Committers team: [William Welling](#)
 - a. Working with TaMU VIVO work
 2. [i18n Sprint updates](#)
 - a. Mini-sprint focused on VIVO committer review/refactor: July 14-16th
 - i. Committers will review work from previous sprint
 1. Benjamin will help with code review
 2. Don has also reserved time on the 14th and 17th
 3. Michel will be available during this time if anyone has questions or needs support
 - ii. Michel: Plans on having good Selenium tests before 14th. Need to make these tests for non-i18n and i18n versions of VIVO and comparing front-ends to ensure that they are exactly the same for the English language version. Ensure some part of the code can be merged into master
 1. <https://wiki.lyrasis.org/display/VIVO/vivo-regression-test%3A+a+Test+Bench+Tool+for+the+Continuous+Evaluation+of+VIVO%27s+Development>
 - iii. Another sprint 2: 24th
 - iv. Benjamin: pull request for Mandarin - wants to do that once i18n branch has been merged into master
 - b. Meeting last week with Andrew, Dominique, and Matthias (and Michel)
 - i. Discussed the sprint
 - ii. One important goal fixed: making the merge to the master branch. Not a way to prove that it is stable enough to put in master branch.
3. Moving priorities forward: Data ingest
 - a. What are the use cases?
 - b. What the "entities" to be ingested? ([reference](#)) ...let's start simple
 - c. Thoughts?
 - i. Don: Using old VIVO harvester and SPARQL program going from CSV -> constructs to get data into target database
 1. Large amounts of data appear to hang the system. Tried various possibilities
 2. Prevents them from upgrading the system
 3. Might be the only institution using the harvester
 4. Seems like every institution is doing their own brand of updates
 5. Appropriate thing: map data to triples, and then input through SPARQL Update API, with indexer and inferencer turned off
 6. VIVO pump? Don't appear to be production sites using this.
 7. Mike suggested RML Mapper. Looking at that. Need rule-based mechanism to map data into triples.
 8. YARRML -<https://rml.io/yarrml/matey/#> . Kent. YAML version that uses rules to mint subjects and objects. Reads from databases, JSON, triples.
 - a. Example:

prefixes:

vitro: "<http://vitro.mannlib.cornell.edu/ns/vitro/0.7/>"

core: "<http://vivoweb.org/ontology/core/>"

vlocal: "<https://experts.colorado.edu/ontology/vivo-fis/>"

vcard: "<http://www.w3.org/2006/vcard/ns#>"

arg: "<http://purl.obolibrary.org/obo/>"

cub: "<https://experts.colorado.edu/individual/>"

mappings:

person:

sources:

- ['fisperson.csv~csv']

s: cub:fisid_\$(FISID)

po:

- [a, foaf:Person]

- [a, foaf:Agent]

- [rdfs:label, \$(LABEL)]

- [ex:name, \$(FIRSTNAME)]

- p: arg:ARG_2000028

o:

- mapping: vcard

condition:

function: equal

parameters:

- [str1, \$(FISID)]

- [str2, \$(FISID)]

vcard:

sources:

- ['fisperson.csv~csv']

s: cub:vcard_\$(FISID)

po:

- [rdf:type, vcard:Kind]

- [rdf:type, arg:ARG_2000379]

- [vitro:mostSpecificType, arg:ARG_2000379]

- [vcard:hasEmail, cub:vcard_name_\$(FISID)~iri]

- [vcard:hasName, cub:vcard_name_\$(FISID)]

1. What are other people using?

2. Michel: How robust is YARRML?

3. Don: RML seems more robust. YARRML translates to RML. RML: creates set of rules and applies to input to create triples. Rules written in RDF/turtle.

4. Main objective: simple way to map the data without having to be an ontologist.

i. Potential use cases

1. Data in CSV or other format, need to map into RDF, and use SPARQL Update API

a. Huda: Anyone remember what RIALTO did and their issues? (We can look up documentation)

b. Don: SPARQL Update API: performance issues. TDB Loader: VM used for VIVO while separate need for tdb loader so problematic.

- c. Don: Want to turn on a switch and upload all the triples. Named graph for every class type. Then turn on inferencer and indexer.
 - i. Don't have real-time use case, so turn off one VIVO instance and use another one for upload.
 - d. Benjamin: Brian worked on a service for turning off indexing and inference for data ingest. Code may be available on DTU repository
 - i. Brian's indexing and inferencing controller service: <https://github.com/RAP-research-output-impact/rap-custom-vivo/blob/69afce74333405e532cd348a7758d690e91fdec3/custom-vivo/webapp/src/main/java/dk/dtu/adm/rap/controller/IndexingInferenceService.java>
 - 2. Michel: Consider VIVO as enterprise data source like other sources inside the organization.
 - a. Need to have messaging system where we can communicate data/exchange data from different data sources
 - b. Evaluating Kafka as a system to do so
 - c. Serialization/mapping data from CSV to RDF: will need to address this in the consumer/updater in the Kafka. Want streaming communication from data sources to VIVO
 - i. (Don from chat: rdfstreamer seems to work with kafka: <https://github.com/RMLio/RMLStreamer>)
 - d. Architecture/sources discussed in VIVO conference see: <http://doi.org/10.13140/RG.2.2.22501.83681>
 - 3. Different use cases with common need for mapping CSV or other data sources to RDF.
 - a. Michel: Don's use case more ETL with VIVO at the center. In our case, integrate across data sources including VIVO.
 - b. Don: Open to other ways of bringing in data that doesn't have to be traditional batch-driven ETL. Streaming messages could serve the same purpose. As long as have enough controls to determine how the app behaves (e.g. how many triples sent over).
 - c. Don: Other paradigm: don't track differences in data. Every class group in own named graph. Easy enough to truncate and reload everything.
 - d. Benjamin: Prefer to upload/edit data through the interface
 - ii. Rachid: Is there a task force for ingest?
 - 1. Huda: Good idea to have one, to gather use cases, determine common requirements, and work on common goals.
 - 2. Rachid: No task force yet or interest group yet
 - iii. RDF messaging
 - 1. William: Approach is tied to Apache Jena, not agnostic. Standard JMS messaging with Artemis broker - significantly different than Kafka stream. Not interchangeable. Similar in goal.
 - iv. To do: bring up interest group/task force idea when Andrew is back
1. [2020-07-15 - Special Topic - VIVO Scholar Next Steps](#)
 - a. Several people on the call planning on attending
 - b. Potential areas of interest
 - i. Robust GraphQL specs
 - ii. Customizable indexes
 - 1. Similar perhaps to the use of LDPATH in Fedora. LDPATH traversal for making configurable solr documents
 - 2. <https://marmotta.apache.org/ldpath/language.html>
 - 3. Compound documents are challenging and mapping to Solr documents from them is challenging
 2. Benjamin: Would be nice to get comments on Andrews PR here: <https://github.com/vivo-project/Vitro/pull/169>
 - a. Better way to load only the language RDF you will be using
 - b. Rearranges the RDF directory structure
 - i. Would put all language files underneath RDF
 - ii. Non-specific language files in directory named core
 - c. Q: Is this better or this worse?
 - d. If people have comments, please feel free to add

Actions

•