

2019-03-26 - VIVO Development IG

Date

26 Mar 2019

Call-in Information

Time: 11:00 am, Eastern Time (New York, GMT-05:00)

To join the online meeting:

- Go to: <https://duraspace.zoom.us/j/823948749>
- Or iPhone one-tap :
 - US: +14086380968,,823948749# or +16468769923,,823948749#
- Or Telephone:
 - Dial(for higher quality, dial a number based on your current location):
 - US: +1 408 638 0968 or +1 646 876 9923 or +1 669 900 6833
 - Meeting ID: 823 948 749
- International numbers available: https://duraspace.zoom.us/join?m=Qy8de-kt6W4fMMDQCAV_3qfH1W-lxAo5

Slack

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

Attendees

[blocked URL](#) Indicating note-taker

1. [Don Elsborg blocked URL](#)
2. [Mike Conlon](#)
3. [Benjamin Gross](#)
4. [Jim Blake](#)
5. [Harry Thakkar](#)
6. Felipe A. Bergallo Corral
7. [Richard Outten](#)
8. [Huda Khan](#)

Agenda

1. Next Week: [When Androids Dream](#) (codename: androdies) - [Huda Khan](#) ([PPT](#) version) (Relegated to next week but feel free to review/give feedback on slides)
2. Search index: approaches for configuration and processes
3. [Sprint update](#)
4. Acceptance tests moved to vivo-community, thanks [Jim Blake](#)
 - a. <https://github.com/vivo-community/vivo-acceptance-tests>
 - b. How to run it?
5. Mailing list messages
 - a. [Some triples not getting deleted in VIVO?](#)
 - b. [How can I backup the custom Pages that have been created?](#)
 - c. [ORCiD Integration - Sync](#) - [Muhammad Javed](#) suggesting [Jim Blake](#) has clues
 - d. [input data Vivo last version](#)
6. 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.

- a. Soft balls

b. Regular t	i.	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.
	ii.	Unable to locate Jira server for this macro. It may be due to Application Link configuration.
	iii.	Unable to locate Jira server for this macro. It may be due to Application Link configuration.
	iv.	Unable to locate Jira server for this macro. It may be due to Application Link configuration.
	v.	Unable to locate Jira server for this macro. It may be due to Application Link configuration.
	vi.	Unable to locate Jira server for this macro. It may be due to Application Link configuration.
	vii.	Unable to locate Jira server for this macro. It may be due to Application Link configuration.
	viii.	Unable to locate Jira server for this macro. It may be due to Application Link configuration.

7. Received

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

8. Bugs (1.1.1)

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

Recording

- <http://bit.ly/vivo-fly-in-design-2019-03>

- Harry created a JIRA ticket to document what he had done with Docker. JIRA-1685.
- Filipe from Sigma, new to VIVO on the call

Huda's presentation

She was tasked with taking notes from the arch flyin. So this is her attempt at that. Requests people who attended fly in to add to notes.

People who want self contained vivo, others who want various components to work better and decouple components. So underlying question of how to decouple but make a better VIVO.

Left side is shapes, rdf, batch updates

Shapes - both for validating information and working with the indexing process.

The UI highlights things for end users but shapes translates that back to something concrete.

The bi-directional INDEX arrow really just means that the Indexing process is reading from the sparql api.

With UI Display, the auth/perms layer is there as a nod that this is what's currently available with VIVO. Not vetted in with Product evolution yet.

Jim - way indexing is done now is that all info in index is public.

Richard - VIVO is still the interface to update info. Graphql only for view.

Mike - Any layered perms for graphql?

Richard - no plan for perms, this is for public display only.

Huda - so the auth/perm box between display and graphql/index is there as a placeholder, nobody is beholden to this.

Next - Combine slide - ETL - Huda is just going off of what people spoke of at fly in.

Seems like validation steps can be in several locations, after mapping, after disambiguation

Mike - elaborate on shapes.

Jim - shapes might not be a hard constraint on ingest, because we don't just want to say all publications MUST have an author. But shapes can be used to validate the data to see if it's optimal.

Huda - would be great to know what is actually in the store.

Don - would shapes be used in Validate - eg show all orphan pubs

Jim - thinks it would be used at ingest to prevent bad data from coming in.

Mike - Can use this to check if IRI's are bad. Perhaps like a mini-graph for ETL

Next slide - of combine

Changes sets

Mike - distinguishing between first class objects (people,pubs, grants) vs others like relationships

Huda - perhaps a separate set that contains relationships.

Don - note that the data loads are done asynchronously. So we might not be able to validate the graph as it's being loaded.

Richard - We determine our URIs for each object in a way that they are known to other objects. So the relationship URI can be know when loading another object.

Next slide - reasoning process

When to trigger reasoning - based on some event?

What is happening now is that when triple is added or removed, reasoning processes are triggered. Pro - things are kept in sync. Con - could be expensive at the wrong time.

Eg - add a faculty member, with class hierarchy we infer they're people, mammals, etc

So danger of the nested russian doll system.

Mike - how many reasoners

Don - abox and tbox, both are used - Jim agreed

Mike - so we need more of this. So paris is a city, paris is in EU, we should query all cities in EU. We need to think much more about this.

Jim - VIVO had to strike a balance between performance and functionality with this.

Mike - So we should be able to get this information on demand, and not have to store the info per-se.

Jim - this wasn't discussed at fly in. We should be able to configure the triple store and customize it.

Huda -Indexing process

Want to extract info from triple store - then create documents from that data.

Can shapes be used here. If a property changes does whole index need to be updated or perhaps just a portion of it. To some extent VIVO is doing this now. So moving forward based on a shape we can create new docs based on changes

Huda - indexing process slide 2

Everything is in place now.

Event configuration of when it happens. On request or when triples are added.

What goes into the index. Mapping.

Here are properties, here are queries, how do I map it to a target document.

What that mapping is it should be configurable. Using documentModifiers we can do this mapping.

Using abstract index we can implement different indexes like SOLR/Elastic/etc

Jim - we talked about modifying indexer to modify multiple indexes - eg one for traditional VIVO index and one used for GraphQL. When talking with Huda and Brian Lowe, maybe this isn't the role. So do we want to shoehorn this into the existing configuration? Perhaps some data desired for an index might come from another source - eg RDBMS

Huda - shapes could inform indexing process.

Slide - Externalization

We have the triplestore and we have multiple ways of asking and receiving things from it. All of this goes through some abstraction - like sparqlAPI.

Actions

- Unable to locate Jira server for this macro. It may be due to Application Link configuration. - Mike Conlon, can you give this one a

Previous Actions

-