

2019-02-06 Brown Dog Meeting notes

Date

06 Feb 2019

Attendees

- Yan Zhao
- Bing Zhang - absent
- Sandeep Puthanveetil Satheesan
- Luigi Marini
- Shannon Bradley -
- Rob Kooper - absent
- Kenton McHenry - absent
- Gregory Jansen - added update
- Michelle Pitcel
- Diego Calderon

Agenda

- Updates - LDAP Yan Zhao following up with Nathan about error with LDAP submission
 - Yan making changes recommended by Rob
 - Rob will need to run it at the end
- CZO use case development
 - Confluence page: [CZO: Geostreaming Data Framework Integration](#)
 - [Copy of Flux Tower New Variable Names](#)
 - Review design with Michelle and Diego to get input
 - Need: if czo gives a csv file - can brown dog determine what the column headers mean
 - Show what we have done so far
 - What can be identified that we can do next?
 - You can do at file level or dataset level
 - does it break if there are files that have different headers within the dataset - yes right now
 - Similar to what Michelle has in parser files - but now it can be done in clowder - user creates the mappings
 - Then there is a service that stores mappings - i.e. temp means this
 - Registered only with csv files - but we could register .dat as a mime type - can be opened as a .csv
 - Will have to figure out how to use this to update old data points - submit for extraction
 - <https://Ameriflux.lbl.gov/data/aboutdata/data-variables/>
 - Share table with Brown Dog team Flux Tower New Variables Names from IMLCZO
 - placed under the CZO planning page: [Copy of Flux Tower New Variable Names](#)
- Tool Catalog Updates/Cleanup
 - Design Page - [Tools Catalog Version 2.0](#)
 - Sandeep put in new wireframes today
- Deployments
 - what about the ability to recreate a container in the correct status - log of what its status was? - need to follow up
 - maintain a status of all extractors that were running so you can bring them back up
- V 0.6
 - <https://opensource.ncsa.illinois.edu/jira/projects/BD/versions/17194>
 - I don't think we are capturing all tasks when we create them in the appropriate release version

Discussion items

Time	Item	Who	Notes
		Shannon	Meeting with Marketing to do blog for Box Skills release Invited Ben and Sandeep to meeting so they can get good links and technical info Due mid week next week Per Isaac G: Illinois should be able to support applying the Brown Dog skill to content in our U of I Box enterprise as soon as it's available. We're coming up with documentation on how users in our enterprise can request a folder that integrates with the Brown Dog skill. As this becomes widely available, I would share this information with the other members of the Box Higher Education Advisory Board so they can integrate Brown Dog in their environments if they choose to. Box added skills to the developer console. @Sandeep you should be able to create one from your Illinois.edu box developer dashboard
		Luigi	

		Yan	LDAP, Pecan then work on Luigi's request ...  GATS-802 - fix swagger.yml DONE
		Sandeep	Wireframes for tools catalog PR for dockerization of fetcher service Which of these need prioritized: Get Galaxy Extractor working Prioritize these, port to the new Pycldowder and deploy to Tools Catalog • Langid • DBPedia • Census From Cell • Handwritten Decimals • Killed Photos • Mean Grey • Faces • Eyes • Profiles • Closeups • NLTK Summary • Stanford CoreNLP • Tesseract • Tika • Versus • VLFeat • Generalized exif/image metadata extractor
		Greg	I ran into a bug trying to get my existing BD tests to run on the new test platform. Was just trying again prior to the meeting, but I got bogged down in it and missed the meeting! Anyway, I have an issue that I'm sorting out. It has to do with my index of test files and the way I am querying it. The site is up, but not results yet. You guys can have whatever scale tests you want when I get this working. I've done some local tests with 32 test workers, each running 2000 requests in 100 seconds. So I can blast BD if you really want me to. (At a convenient time on DEV)
		All	